NOAA Technical Memorandum OAR GSD-38

QUALITY ASSESSMENT OF CoSPA

S.A. Lack
G.J. Layne
S. Madine
J.L. Mahoney

Earth System Research Laboratory
Global Systems Division
Boulder, Colorado
August 2011

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Office of Oceanic and Atmospheric Research

NOAA Technical Memorandum OAR GSD-38

QUALITY ASSESSMENT OF CoSPA

Steve A. Lack[2]
Geary J. Layne[2]
Sean Madine[3]
Jennifer L. Mahoney[1]

[1]National Oceanic and Atmospheric Administration, Earth System Research laboratory, Global Systems Division, (NOAA/ESRL/GSD)

[2]Cooperative Institute for Research in Environmental Sciences (CIRES) and NOAA/ESRL/GSD

[3]Cooperative Institute for Research in the Atmosphere (CIRA) and NOAA/ESRL/GSD

Earth System Research Laboratory
Global Systems Division
Boulder, Colorado
August 2011

UNITED STATES DEPARTMENT OF COMMERCE

Dr. Rebecca M. Blank
Acting Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Dr. Jane Lubchenco
Under Secretary for Oceans and Atmosphere/Administrator

Office of Oceanic and Atmospheric Research

Craig McLean
Acting Assistant Administrator

NOTICE

Mention of a commercial company or product does not constitute an endorsement by NOAA/Earth System Research Laboratory. Use of information from this publication concerning proprietary products or the test of such products for publicity or advertising purposes is not authorized.

For sale by the National Technical Information Service, 5285 Port Royal Road
Springfield, VA 22061

Table of Contents

1. Introduction 1

2. Approach 1

2.1 Weather assessment and weather translation assessment 1

2.2 Assessment Methodology 3

2.2.1 Data 3

2.2.2 Methodology Overview 5

2.2.3 Fractions Skill Score 6

2.2.4 Mincut-Bottleneck 7

2.2.5 Verification at CCFP-like scale 11

3. Analysis 14

3.1 Climatology 14

3.2 Forecast performance at the 4 to 6-h lead-time 17

3.2.1 Performance in the northeast domain 17

3.2.2 Performance in the southeast domain 19

3.3 Forecast performance at the 2 to 4-h lead-time 21

3.3.1 Performance in the northeast domain 21

3.3.2 Performance in the southeast domain 21

3.3.3 CoSPA performance for 2 to 4-h lead-times 23

3.4 Performance at the 8-h lead-time 24

3.5 Supplemental relationship to CCFP 25

3.5.1 CCFP sparse coverage/low confidence polygons 25

3.5.2 CoSPA/CCFP relationship 25

3.5.3 Long-range planning 25

3.6 Blitz Days versus Full Season Skill 26

3.6.1 Climatology of Blitz Days 26

3.6.2 Blitz vs. Full Season Skill Comparison 28

4. Conclusions 29

Acknowledgements 30

5. References 30

6. Appendix 32

6.1 FSS with Confidence Intervals 32

6.2 Mincut with Confidence Intervals 33

List of Figures

Figure 2.1: Planning process from the JPDO ATM Weather Integration Plan (Draft v1.5, 30 June 2010). The orange columns represent sequential aspects of planning, with arrows representing the flow from weather forecasts (left) to air traffic management decisions (right). The blue boxes depict the two main areas of the QA PDT assessment (weather and weather translation). Verification techniques, described in the next section, were developed and tuned specifically for each of these two areas. 1

Figure 2.2: Extent of the CoSPA domain (shaded) with northeast (red) and southeast (blue) domains 2

Figure 2.3: Example of the CoSPA VIL forecast field issued at 1500 UTC valid at 2100 UTC on 7 July 2010. 4

Figure 2.4: Example of the LCH product issued at 0000 UTC valid between 0200 UTC and 0400 UTC on 5 May 2010. 5

Figure 2.5: Example of 5x5 neighborhoods about the center pixel for input into the FSS. (After E. Ebert) 7

Figure 2.6: Simplified 10x10 grid showing an example of the Mincut-Bottleneck technique. The deterministic convective objects are shown in red along the implied flow in gray for the given corridor shown in blue. The minimum paths across the corridor are given in parentheses and shown with arrows. 8

Figure 2.7: Simplified 10x10 grid showing an example of the Mincut-Bottleneck technique for a probabilistic field. Shades of red show varied probabilistic convective fields between the corridor bounds shown in blue. The minimum paths across the corridor are given in parentheses and shown with arrows. 9

Figure 2.8: An example of the Mincut-Bottleneck technique for a given hexagon given a convective node (red). The mincut is shown by a green arrow for one corridor (blue) of the hexagonal cell in the given hexagon that results in a flow reduction for the given flow direction (yellow) (left). The resulting blockage for all three corridors in the hexagon is shown (right). 10

Figure 2.9: 75-nm hexagons over CONUS. 11

Figure 2.10: 300-nm hexagons over CONUS. 11

Figure 2.11: Scaling example for a particular convective event in Texas and Oklahoma. Actual reflectivity (top), with the output from the clustering algorithm (bottom) in arbitrary units, which can then be converted to areas assigned to coverage categories. 12

Figure 2.12: An example of a scaled CIWS field valid at 2100 UTC on 13 June 2010. Light green represents sparse/low coverage, dark green sparse/high coverage, yellow medium coverage or higher, and red is actual CIWS VIL at 3.5 kg m^{-2} or greater. 13

Figure 3.1: Convective coverage (VIP-level 3 or greater) by hour of day for the northeast (top) and the southeast (bottom). The solid curves represent the coverage of the observed weather (right y-axis), while the dotted lines represent bias for the different products at the 6-h lead-time (left y-axis). Bias is given on a log base 2 scale so that distance away from 1 is shown as equal penalty for over- or under-forecasting. At the time of strategic planning, both curves are near minimum. 14

Figure 3.2: Normalized climatology of CIWS VIL analysis at a threshold of VIP-level 3 valid at 2100 UTC each day from 1 June 2010 through 30 September 2010. Larger values (darker colors) indicate higher occurrence of convective weather. 15

Figure 3.3: Normalized climatology of CoSPA forecast (1500 UTC issuance, 6-h lead-time) at a threshold of VIP-level 3 valid at 2100 UTC each day from 1 June 2010 through 30 September 2010. Larger values (darker colors) indicate higher occurrence of forecast convection. 15

Figure 3.4: Normalized climatology of the LAMP Thunderstorm Product (1500 UTC issuance, 6-h lead-time) valid at 2100 UTC each day from 1 June 2010 through 30 September 2010. Larger values (darker colors) indicate higher occurrence of forecast convection. 16

Figure 3.5: Normalized climatology of calibrated CCFP (1500 UTC issuance, 6-h lead-time) valid at 2100 UTC each day from 1 June 2010 through 30 September 2010. Larger values (darker colors) indicate higher occurrence of forecast convection. 16

Figure 3.6: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), LAMP (purple), climo (black), and uniform (grey). Scores include northeast domain only, for products issued at 1500 UTC with a 6-h lead-time. Note: resolution is not a linear scale. 17

Figure 3.7: Critical Success Index (CSI) versus Mincut threshold for ARTCC scale (left) and sector scale (right). Products shown are CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), and LAMP (purple). The dashed grey line indicates the mean number of hexagons that are impacted at the given Mincut threshold for all days in the observed field, with the scale on the right of the plot. Vertical dashed lines represent medium impact events (yellow) and high impact events (maroon). Scores include northeast domain only, for products issued at 1500 UTC with a 6-h lead-time. 18

Figure 3.8: Bias versus Mincut threshold for ARTCC scale (left) and sector scale (right). Products shown are CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP(cyan), and LAMP (purple). The dashed grey line indicates the mean number of hexagons that are impacted at the given Mincut threshold for all days in the observed field, with the scale on the right of the plot. Vertical dashed lines represent medium impact events (yellow) and high impact events (maroon). Scores include northeast domain only, for products issued at 1500 UTC with a 6-h lead-time. 18

Figure 3.9: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), LAMP (purple), climo (black), and uniform (grey). Scores include southeast domain only, for products issued at 1500 UTC with a 6-h lead-time. Note: resolution is not a linear scale. 19

Figure 3.10: Critical Success Index (CSI) versus Mincut threshold for ARTCC scale (left) and sector scale (right). Products shown are CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), and LAMP (purple). The dashed grey line indicates the mean number of hexagons that are impacted at the given Mincut threshold for all days in the observed field, with the scale on the right of the plot. Vertical dashed lines represent medium impact events (yellow) and high impact events (maroon). Scores include southeast domain only, for products issued at 1500 UTC with a 6-h lead-time. 20

Figure 3.11: Bias versus Mincut threshold for ARTCC scale (left) and sector scale (right). Products shown are CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), and

LAMP (purple). The dashed grey line indicates the mean number of hexagons that are impacted at the given Mincut threshold for all days in the observed field, with the scale on the right of the plot. Vertical dashed lines represent medium impact events (yellow) and high impact events (maroon). Scores include southeast domain only, for products issued at 1500 UTC with a 6-h lead-time. 20

Figure 3.12: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), LAMP (purple), climo (black), and uniform (grey). Scores include northeast domain only, for products issued at 1900 UTC with a 2-h lead-time. Note: resolution is not a linear scale. 21

Figure 3.13: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), LAMP (purple), climo (black), and uniform (grey). Scores include southeast domain only, for products issued at 1900 UTC with a 2-h lead-time. Note: resolution is not a linear scale. 22

Figure 3.14: Convective coverage (VIP-level 3 or greater) by hour of day for the southeast domain. The solid curve represents the coverage of the observed weather (right y-axis), while the dotted lines represent bias for the different products at the 2-h lead-time (left y-axis). Bias is given on a log base 2 scale so that distance away from 1 is shown as equal penalty for over- or under-forecasting..... 22

Figure 3.15: Fractions Skill Score versus resolution for CoSPA lead times between two and four hours in the northeast domain (top) and the southeast domain (bottom), for products issued at 1500 UTC. Note: resolution is not a linear scale..... 23

Figure 3.16: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), LAMP (purple), climo (black) and uniform (grey) in the northeast domain (top) and the southeast domain (bottom), for products issued at 1300 UTC with an 8-h lead-time. Note: resolution is not a linear scale. 24

Figure 3.17: CSI versus lead-time for LAMP, each forecast lead time valid at 2100 UTC, as determined by the sparse category of the scaling technique. 25

Figure 3.18: CIWS climatology for the blitz days selected for human evaluation (top) compared with the CIWS climatology for the full season (bottom) valid at 2100 UTC..... 27

Figure 3.19: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), LAMP (purple), CCFP (red), calibrated CCFP (cyan), climo (black) and uniform (grey) in the northeast domain for blitz days (top) and for the full season (bottom), for products issued at 1500 UTC with a 6-h lead-time. Note: resolution is not a linear scale. 28

List of Tables

Table 2.1. Observed climatological percent coverage of CIWS in CCFP by polygon type and lead time. 6

Table 2.2. LAMP probability (%) thresholds for the creation of CCFP-like polygons..... 13

Executive Summary

This report summarizes a formal quality assessment of CoSPA, a high-resolution convective forecast being developed for use in air traffic management. On behalf of the Federal Aviation Administration's Aviation Weather Research Program, and in support of an Aviation Weather Technology Transfer (AWTT) D4 (operational) decision point, this study was carried out during the 2010 convective season as outlined in the FAA Aviation Weather Group's *Comprehensive Plan for the CoSPA 2010 Demonstration*.

The Quality Assessment Product Development Team (QA PDT) worked to determine the value of this forecast for the 2 to 8-h lead-time, as it pertains to strategic Traffic Flow Management (TFM). The current operational forecast, the Collaborative Convective Forecast Product (CCFP), was used to define the baseline of performance against which the experimental product was compared. Results for the Localized Aviation MOS Program/CCFP Hybrid (LAMP CCFP Hybrid, LCH), another product used in the 2010 evaluation, are also included in this assessment.

The assessment framework for this report adopts concepts from the Joint Planning and Development Office (JPDO) Air Traffic Management (ATM) Weather Integration Plan as a description of the planning process, whereas the day-to-day air traffic planning process is much more difficult to define. To determine indicators of forecast value, the QA PDT assessment focused on two areas of investigation within the planning process: *weather* and *weather translation*. As the baseline forecast for the study, CCFP was treated in our analysis in two different ways: by straightforward definition, and calibrated based on its operational performance over recent years.

Primary themes in the Weather Assessment:

- Direct analysis of CCFP, CoSPA, and LCH in the context of strategic planning; including stratifications for planning times, with emphasis on issuances before convective initiation.
- Comparison of CoSPA and the LAMP Thunderstorm Product as supplements to CCFP; forecast relationship outlined in the concept of use for the 2010 Operational Evaluation.
- Diagnostic comparisons of CoSPA and the underlying High-Resolution Rapid Refresh (HRRR) model forecasts.

Primary themes in the Weather Translation Assessment:

- Direct analysis of CCFP, CoSPA, and LCH in the context of strategic planning; including stratifications for planning times, with emphasis on issuances before convective initiation.
- Diagnostic comparisons of CoSPA and the underlying HRRR model forecasts.

Climatology

This analysis determined that convective weather regimes differ significantly in the northeast and southeast U.S., with corresponding delineation in forecast quality in these two domains.

Performance at the 4 to 6-h lead-time

In the northeast, CoSPA outperforms the other forecasts, showing good skill on a scale similar to that of high-altitude sectors used for TFM. Weather translation metrics indicate that CoSPA forecasts more intense, rare events (sharpness), while LAMP does not.

In the southeast, weather metrics indicate LAMP and calibrated CCFP perform significantly better than CoSPA, but only at relatively coarse scales. Weather translation metrics indicate that LAMP only forecasts lower-intensity events at ARTCC-scale, while calibrated CCFP appears to provide information about the higher intensity ARTCC-scale events.

Performance at the 2 to 4-h lead-time

In the northeast and southeast, both weather and weather translation metrics indicate that CoSPA performs as well as or better than the other forecasts. CoSPA does exhibit a visible discontinuity between the 2:00 (actually the CIWS 2:00) and the 2:15 leads, likely due to the difference between CIWS and the underlying blended forecast. Beyond that time, CoSPA forecasts appear to be consistent.

Performance at the 8-h lead-time

Overall, the 8-h lead-time forecasts of CoSPA and LAMP perform similar to their 6-h lead-time counterparts. The observations about the products at the 6-h lead-time pertain also to the 8-h lead-time. Thus, the 8-h lead-time forecasts appear to have planning value similar to that of the 6-h forecasts.

Supplemental relationship to CCFP

Analysis based on the use of the forecasts in conjunction with one another indicates:

- CoSPA adds information to CCFP sparse coverage, low confidence polygons issued at strategic planning times, by effectively discriminating those that will contain significant areas of convection.
- CoSPA performs much better inside CCFP polygons than in areas outside these polygons. Planners might use the finer structure provided by CoSPA inside the polygons with a greater degree of confidence.
- LAMP provides useful information for next-day planning (12 and 24-h lead-times) for convective coverage at the broadest scale.

QA PDT Study Relative to Other 2010 Operational Evaluation Studies

The climatology of "blitz days", which are typically strongly forced scenarios, is significantly different from that of the overall study period. Differences between the QA and other studies could be the result of differently sampled meteorological distributions.

1. Introduction

This report summarizes a formal quality assessment of CoSPA, a high-resolution convective forecast being developed for use in air traffic management. The QA PDT worked to determine the value of this forecast for the 2 to 8-h lead-time, as it pertains to strategic Traffic Flow Management (TFM). The current operational convective forecast, CCFP, was used to define a baseline of performance against which the experimental product was compared. Results for the LCH, another product used in the 2010 evaluation, are also included in this assessment.

2. Approach

2.1 Weather assessment and weather translation assessment

The assessment framework for this report adopts concepts from the Joint Planning and Development Office Air Traffic Management Weather Integration Plan as a description of the planning process, whereas the day-to-day air traffic planning process is much more difficult to define. **Figure 2.1** depicts the flow of weather information in the context of air traffic management, as defined by the JPDO ATM Weather Integration Team. To determine indicators of forecast value, the QA Assessment focused on two areas of investigation within the planning process: *weather* and *weather translation* (see caption for **Figure 2.1**).

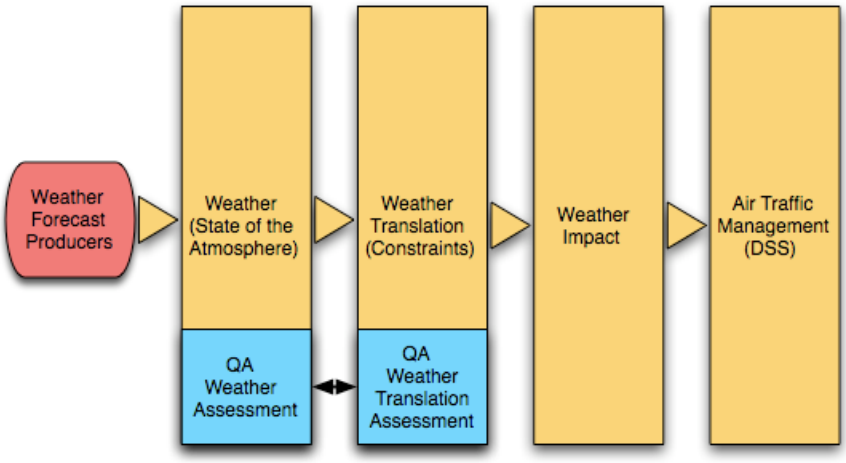

Figure 2.1: Planning process from the JPDO ATM Weather Integration Plan (Draft v1.5, 30 June 2010). The orange columns represent sequential aspects of planning, with arrows representing the flow from weather forecasts (left) to air traffic management decisions (right). The blue boxes depict the two main areas of the QA PDT assessment (weather and weather translation). Verification techniques, described in the next section, were developed and tuned specifically for each of these two areas.

Primary themes in the Weather Assessment:

- Direct analysis of CCFP, CoSPA, and LCH in the context of strategic planning; including stratifications for planning times, with emphasis on issuances before convective initiation.
- Comparison of CoSPA and the LAMP Thunderstorm Product as supplements to CCFP; forecast relationship outlined in the Concept of Use for the Operational Evaluation.
- Diagnostic comparisons of CoSPA and the underlying High-Resolution Rapid Refresh (HRRR) model forecasts.

Primary themes in the Weather Translation Assessment:

- Direct analysis of CCFP, CoSPA, and LCH in the context of strategic planning; including stratifications for planning times, with emphasis on issuances before convective initiation.
- Diagnostic comparisons of CoSPA and the underlying HRRR model forecasts.

As the baseline forecast for the study, CCFP was treated in our analysis in two different ways: by straightforward definition, and calibrated based on its operational performance over recent years. The calibrated CCFP analysis indicates the performance of the forecast as it would be viewed by an experienced user of the product, an individual who would naturally perform a "human calibration" during planning.

Earlier QA PDT studies of CoSPA and LCH found that the products performed differently in the northeast and southeast regions of the CONUS. Convection in the northeast is typically associated with strong frontal forcing, while in the southeast it is more dominated by diurnal heating within a large air mass. The present analysis stratifies all results by geographic region (as shown in **Figure 2.2**) because the difference in performance for these regions is so pronounced.

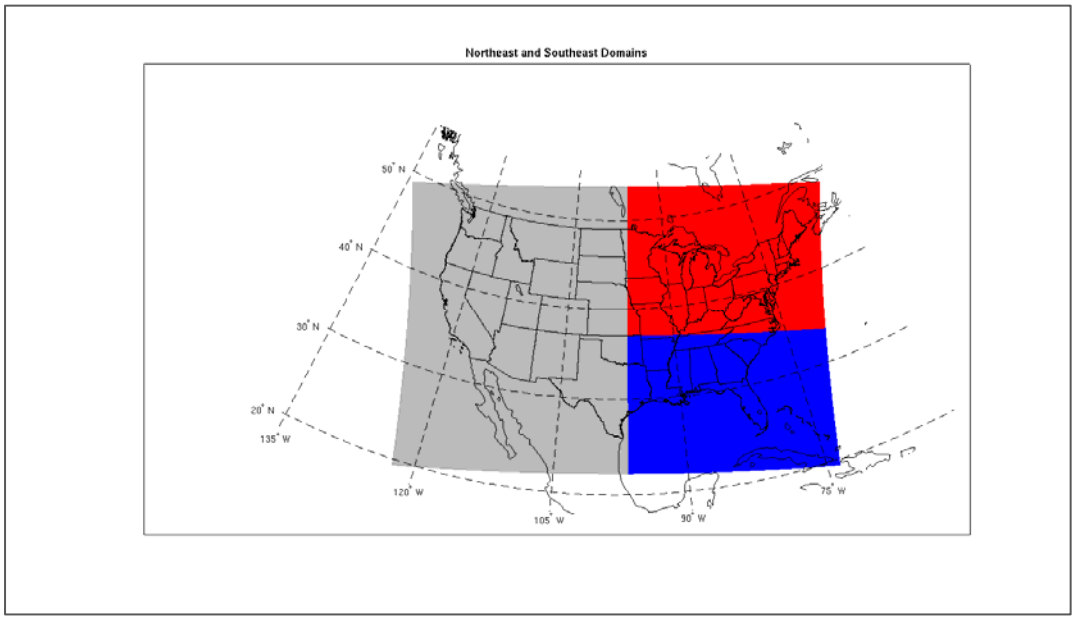

Figure 2.2: Extent of the CoSPA domain (shaded) with northeast (red) and southeast (blue) domains.

2.2 Assessment Methodology

This section describes the framework and techniques used in the assessment to address themes from the context of *weather* and *weather translation*:

- Data and methodology overview
- Fractions Skill Score, as a primary measure regarding weather forecast quality
- Mincut-Bottleneck technique, as a primary measure regarding weather translation
- Forecast scaling, for use in the supplemental analysis

2.2.1 Data

Current air traffic operations use convective weather information to issue airspace flow programs (AFP), ground delay programs (GDP), and other route advisories. As the national airspace becomes increasingly saturated, additional forecasts that attempt to provide more structural information are being evaluated to supplement the current operational baseline, the CCFP. Both forecasts that describe deterministic structure (e.g. simulated radar reflectivity) and probabilistic occurrence are being considered as supplements to CCFP.

CoSPA, as it was evaluated, is a Contiguous United States (CONUS), 0 to 8-h convective forecast of vertically integrated liquid water (VIL) and echo tops at 1 km resolution with a 15-min update cycle (Wolfson et al. 2007). CoSPA is made up of three main components, including: (1) an extrapolation forecast; (2) a high-resolution numerical weather prediction (NWP) model, and (3) a blending algorithm. The 0 to 2-h CoSPA forecast is simply the extrapolation forecast from the Corridor Integrated Weather System (CIWS). CIWS has a 2.5-min update cycle with 5-min lead-time increments and displays forecasts of VIL and echo top at 1 km resolution (Dupree et al. 2009). The 2 to 8-h CoSPA forecast uses a blending algorithm (Phillips et al. 2010) to combine the CIWS 0 to 2-h extrapolation forecast with the High Resolution Rapid Refresh (HRRR) model output (described by Weygandt et al. 2010). The HRRR model was available hourly with 15-min lead-time increments and contained both VIL and echo top fields out to the 15-h lead-time. As input to the CoSPA blending algorithm, the model typically had a 3-h latency. The display for CoSPA is shown in **Figure 2.3**. CCFP polygons were overlaid in the situational display, as CCFP is still the standard for convective weather forecasts. VIL values shown in the forecast are grey for VIP-level 1 and 2 equivalents, yellow for VIP-level 3 and 4, and red for values that meet or exceed VIP-level 5. Hazardous convection as defined by the FAA is VIP-level 3 and above. The echo top values displayed contain four levels of intensity in shades of purple: light purple for echo tops less than 30 kft, and dark purple for echo tops 40 kft and above at 5-kft intervals.

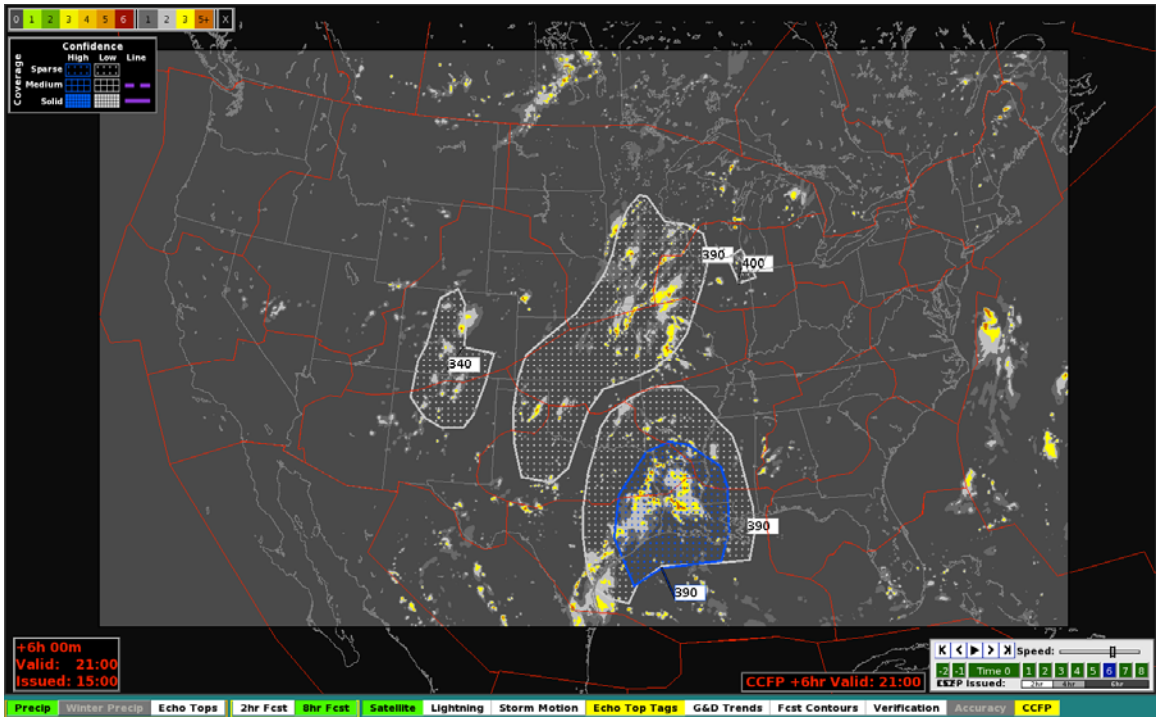

Figure 2.3: Example of the CoSPA VIL forecast field issued at 1500 UTC valid at 2100 UTC on 7 July 2010.

The Localized Aviation MOS (Model Output Statistics) Program (LAMP) is a forecast system that produces post-processed statistical output from the Global Forecast System (GFS) model (Ghirardelli, 2005). The LAMP Thunderstorm Probability field uses recent surface observations combined with the GFS and a climatological background to produce a field of probabilities for the likelihood of a thunderstorm in a 2-h window. The definition of thunderstorm is closely tied with the occurrence of lightning. The LAMP Thunderstorm Probability field is available on the National Weather Service's (NWS) National Digital Forecast Database (NDFD) 5 km grid at 1-h update cycles out to a 25-h lead-time. The LAMP Thunderstorm Probability field is used with current CCFP polygons overlaid to produce a LAMP-CCFP Hybrid (LCH) product. An example of the LCH product is shown in **Figure 2.4**.

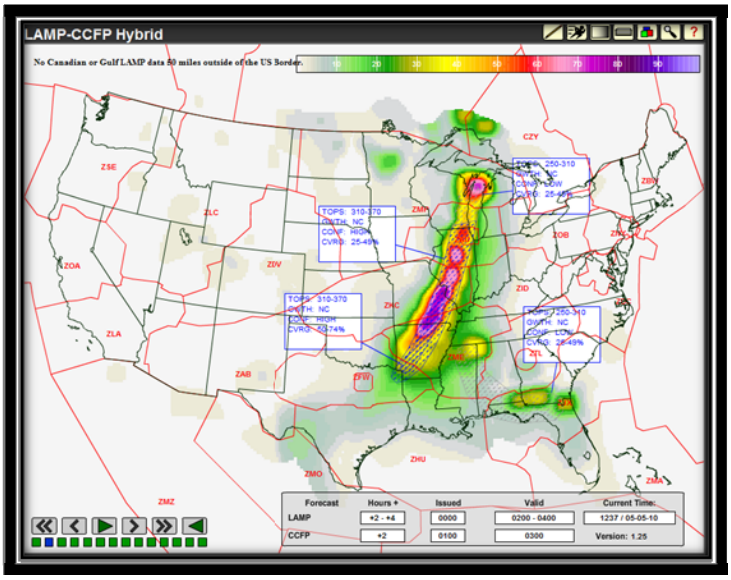

Figure 2.4: Example of the LCH product issued at 0000 UTC valid between 0200 UTC and 0400 UTC on 5 May 2010.

CCFP is the operational convective forecast standard for air traffic operations and serves as the baseline for this study. CCFP was updated for the 2010 convective season. The notable changes include the delineation of medium coverage and solid coverage lines, the redefinition of the sparse coverage bin to 25-40% and medium coverage to 40-74%, the redefinition of echo top bins to 25-29 kft, 30-34 kft, 35-39 kft, and 40 kft and above, and the elimination of the fast growth category.

2.2.2 Methodology Overview

The 2010 summer convective study focused on the evaluation of CoSPA and the LAMP CCFP Hybrid (LCH) product with CCFP as the baseline reference forecast. The High Resolution Rapid Refresh (HRRR) and a calibrated CCFP product were also evaluated as secondary products. The primary observation field was the CIWS analysis field for both VIL and echo tops, with NCWD examined as a secondary measurement. The primary threshold examined pertains to the current definition of hazardous convection, VIP-level 3, which corresponds to a VIL value of 3.5 kg m^{-2}.

Comparisons between CoSPA and its underlying HRRR input account for the latency of the model, which in most cases was 3 hours as indicated by CoSPA file headers. All issuances and leads of these products were examined during the analysis; however, the results presented in this report focus on the strategic telecon times of 1100, 1300, and 1500 UTC and the 2 to 8-h lead-times. An additional examination of longer lead times (12 and 24-h) was performed for the LCH product.

CCFP was evaluated by examining the coverage of the given observation at product definition (VIP-level 3) at a specific time (no window). It is well known that the values of the defined coverage categories for CCFP polygons are significantly higher than those observed climatologically, so a calibrated CCFP was developed. The calibrated CCFP simply changes the

definition of a given coverage/confidence category to the observed climatological value by lead-time. The observed climatology for CCFP polygons by type and lead are shown in **Table 2.1**. Medium coverage and solid (high) coverage were combined into one category due to the rarity of these highest coverage polygons.

Table 2.1. Observed climatological percent coverage of CIWS in CCFP by polygon type and lead time.

	2-h Lead	4-h Lead	6-h Lead
Sparse/Low	4.64	3.93	3.62
Sparse/High	9.77	7.75	6.76
Medium Plus	20.38	14.87	13.15

Qualitatively, the climatological characteristics of the forecast products and the observed weather during the 2010 convective season were examined. The climatological grids were created by averaging the occurrence of convection (observed or forecasted) at each grid box for the set of days used in the study. A Gaussian smoothing operator was then applied to the grid, which retains systematic signals on relevant scales. The grids were then normalized to a common color scale. These images provide insight into the overall location and structure of convection.

Three main quantitative methodologies are described in detail below and make up the bulk of the quantitative assessment.

2.2.3 Fractions Skill Score

The Fractions Skill Score (FSS), described by Roberts and Lean (2008), is a neighborhood verification approach commonly used to assess the skill of numerical weather prediction (NWP) products at various resolutions. The FSS compares the percent coverage of the forecast to the percent coverage of the observations for a given neighborhood about a reference pixel for all pixels in the forecast field. The FSS is given by equation (1), and is defined as the average sum squared difference of the percent coverage in the forecast and observations divided by the average sum of the squares of the percent coverage of the forecast and observations. The FSS has a valid range between 0 (worst) and 1 (best) where values over a defined baseline are said to have skill.

$$FSS = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}\left(P_{fcst} - P_{obs}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}P_{fcst}^{\,2} + \frac{1}{N}\sum_{i=1}^{N}P_{obs}^{\,2}} \qquad (1)$$

Figure 2.5 shows an example of how percent coverage is calculated in a domain. A 5x5 neighborhood is created around the center pixel for this example. The observation at the center pixel would get a P_{obs} value of 0.32 and the forecast at the center pixel would get a P_{fcst} value of 0.44. This procedure is repeated for all pixels in the native domain, and the results are input into equation 1 for the calculation of the FSS over a 5x5 neighborhood.

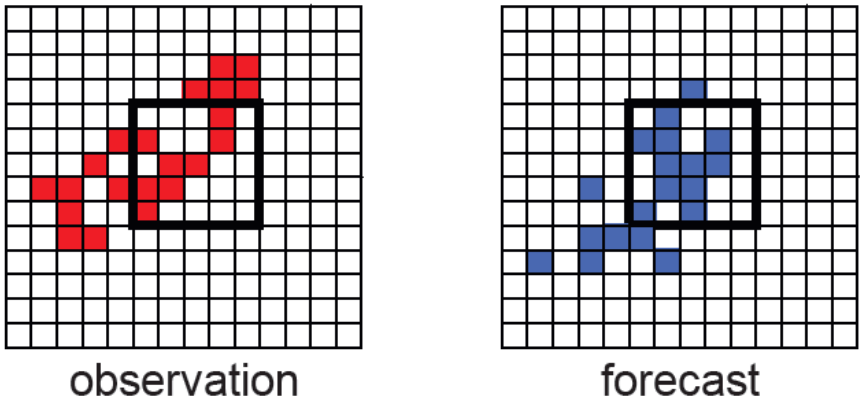

Figure 2.5: Example of 5x5 neighborhoods about the center pixel for input into the FSS. (After E. Ebert)

The FSS can be used to assess the skill of both dichotomous and probabilistic forecasts. In the case of a dichotomous forecast, pixels are assigned a binary value for a given threshold. Scores for probabilistic forecasts are computed with the native pixel values. The CIWS observation field and CoSPA forecast FSS grids were computed using a vertically integrated liquid (VIL) threshold of 3.5 kg m^{-2}. CCFP probabilities use the lower end of the coverage categories in the product definition, and the LAMP thunderstorm probability grid was unmodified for FSS calculations. For this study a circular neighborhood was used with the following radii: 3, 9, 21, 30, 45, 60, 90, 120, 180, 240, 300, and 360 km.

2.2.4 Mincut-Bottleneck

As a primary measure of weather translation quality, the Mincut-Bottleneck technique was applied to examine the structure of convection at various resolutions relevant to air traffic management. The methodology follows closely that of Layne and Lack (2010), which is an approach adapted from Krozel et al. (2004). It yields flow reduction estimates from proposed air space management for NextGen applications, which indirectly serve as a measure of convective structure.

The Mincut-Bottleneck approach estimates permeability by calculating the minimum distance across a specified geometry from source to sink (perpendicular to the corridor of flow) using convective objects as nodes. The flow reduction is defined as one minus the ratio of the minimum distance found with convection present to the minimum distance across the corridor (2). This reduction in flow is calculated and compared for both the forecast and observation.

$$\text{Flow Reduction} = 1 - \frac{Mincut_{convection}}{Mincut_{corridor}} \quad (2)$$

A simplified example using a regular grid for a deterministic forecast or observation is shown in **Figure 2.6**. The method is easily transferable as path lengths (PL) for probabilistic forecasts or observations as shown in **Figure 2.7**. The minimum paths across the corridor are given in parentheses and shown with arrows; the resultant flow reduction is also given.

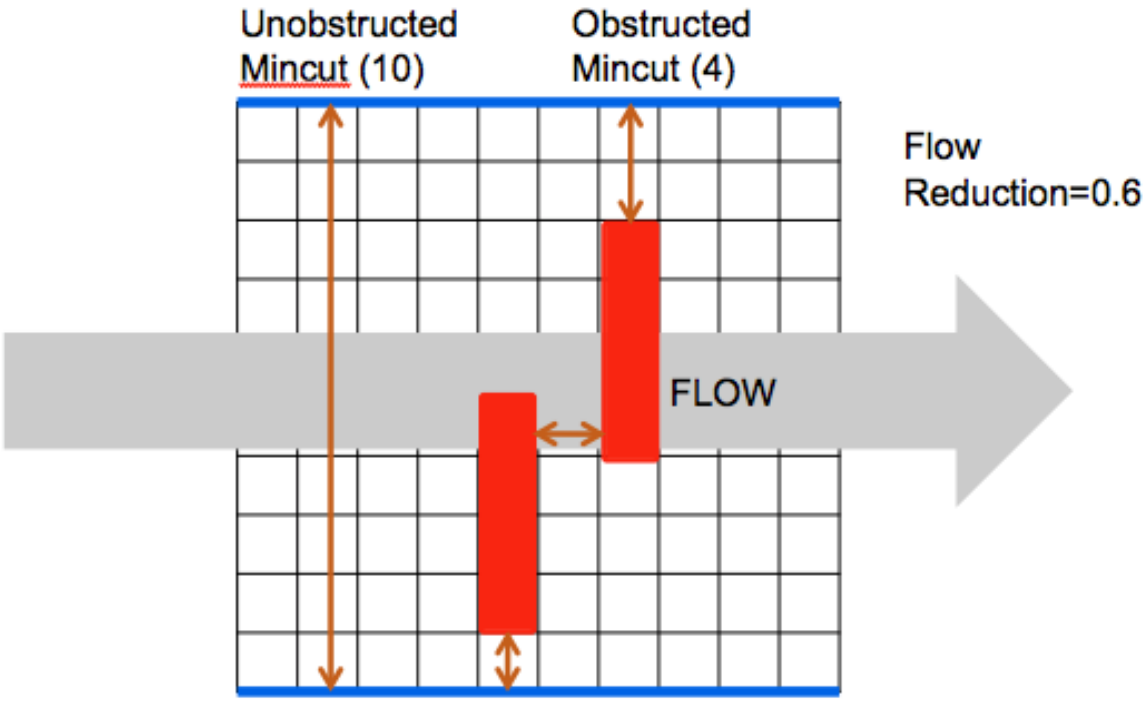

Figure 2.6: Simplified 10x10 grid showing an example of the Mincut-Bottleneck technique. The deterministic convective objects are shown in red along the implied flow in gray for the given corridor shown in blue. The minimum paths across the corridor are given in parentheses and shown with arrows.

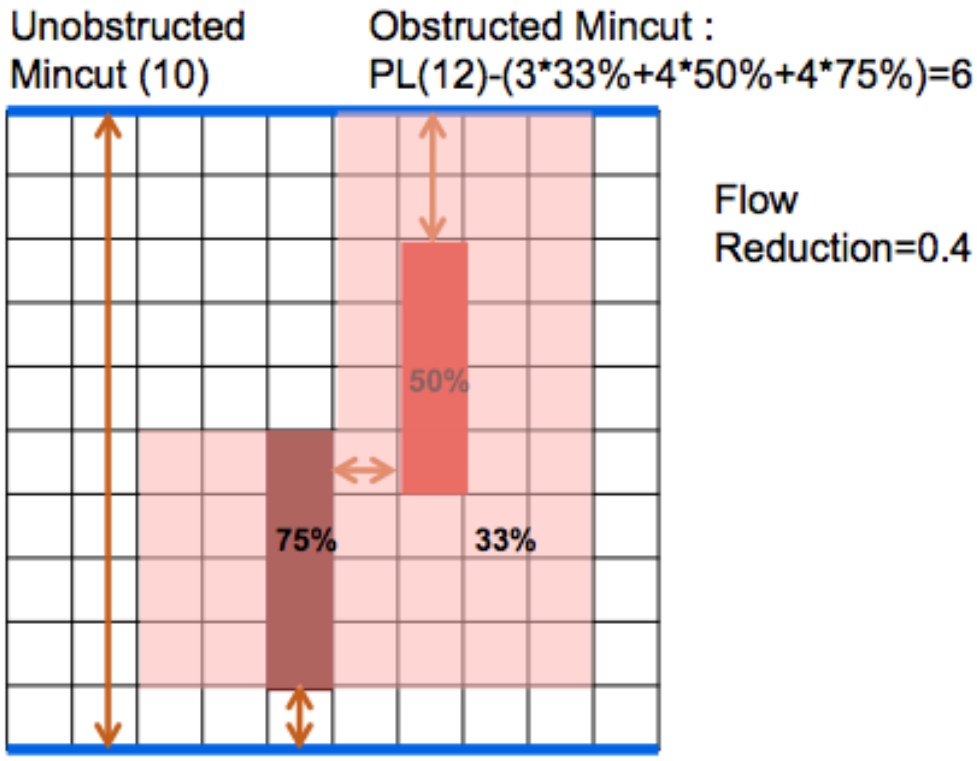

Figure 2.7: Simplified 10x10 grid showing an example of the Mincut-Bottleneck technique for a probabilistic field. Shades of red show varied probabilistic convective fields between the corridor bounds shown in blue. The minimum paths across the corridor are given in parentheses and shown with arrows.

In this study, the Mincut-Bottleneck approach calculates the minimum distance across a hexagonal cell (see **Figure 2.8**) from a source and sink node using convective hazards as nodes. The hexagonal grid cell defines three corridors of flow, each being assigned a Mincut value. **Figure 2.8** (left) shows an example, with one corridor of the hexagonal cell highlighted by the blue lines. In this case, the convective hazard node (in red) is defined by the CIWS analysis VIL at a threshold of 3.5 kg m^{-2}. The green arrow represents the minimum distance from the convective node to the edge of the corridor. For each of the three corridors, the algorithm provides a value to be used in the computation of the metric. The lines at the center of the hexagonal cell in **Figure 2.8** (right), each perpendicular to its corridor of interest, indicate the restriction along that corridor.

In computing Mincut values, the algorithm does not apply a threshold to probabilistic forecasts as in the example above. It computes distances using "fractions obstructions" rather than the complete blockages defined by a dichotomous threshold. Dichotomous forecasts (CoSPA and HRRR) are simply treated as probability of 1 where the threshold of VIP-level 3 is met or exceeded and 0 for areas less than this threshold.

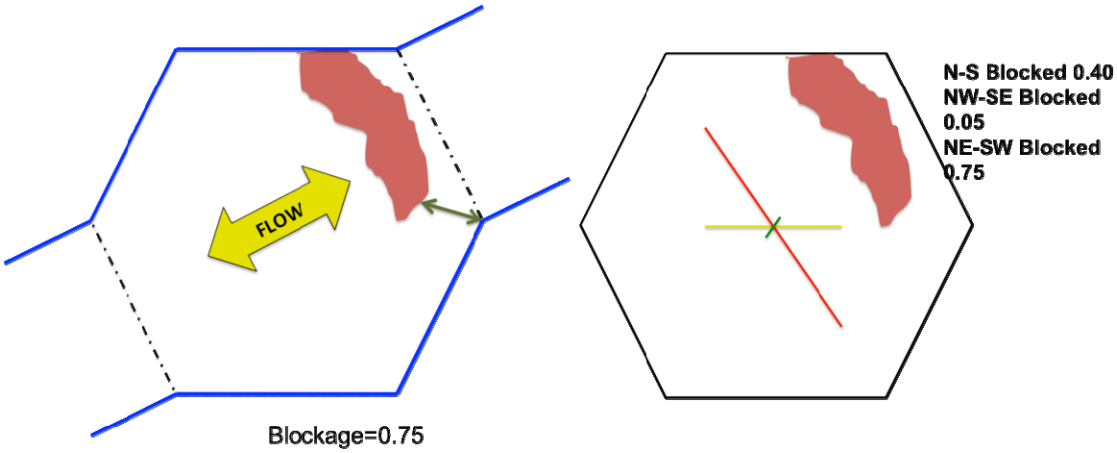

Figure 2.8: An example of the Mincut-Bottleneck technique for a given hexagon given a convective node (red). The mincut is shown by a green arrow for one corridor (blue) of the hexagonal cell in the given hexagon that results in a flow reduction for the given flow direction (yellow) (left). The resulting blockage for all three corridors in the hexagon is shown (right).

For this study, two hexagon heights were chosen: 75 nm and 300 nm. The 75-nm hexagon approximates the size of the average super high sector, while the 300-nm hexagon approximates the size of Air Route Traffic Control Centers (ARTCCs). These hexagonal grids are shown in **Figure 2.9** and **Figure 2.10**, respectively. For this study, all products were calibrated in a way such that the number of impacted hexagons from the observations matches the number of impacted hexagons from the forecast on average for the season.

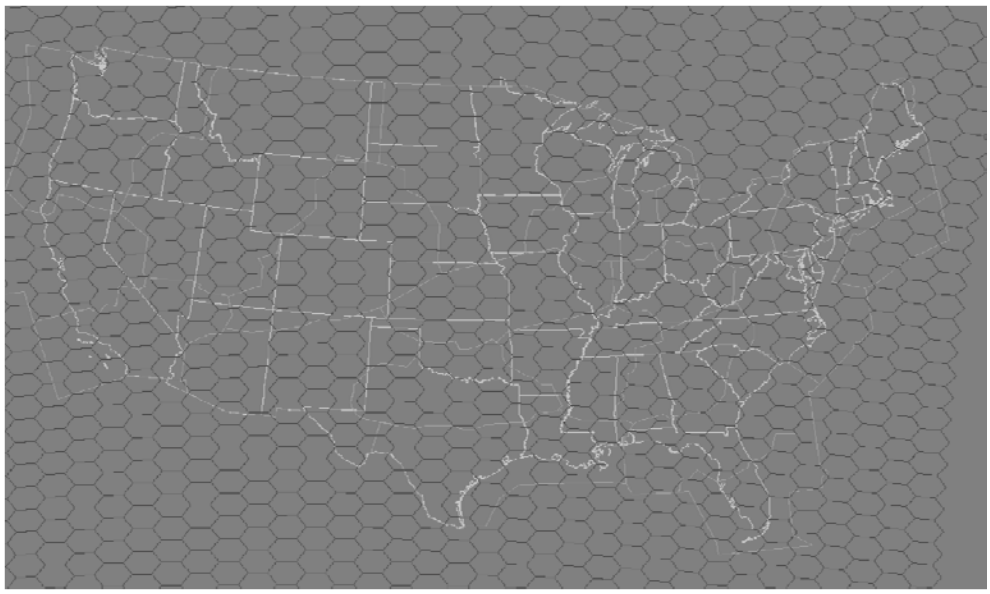

Figure 2.9: 75-nm hexagons over CONUS.

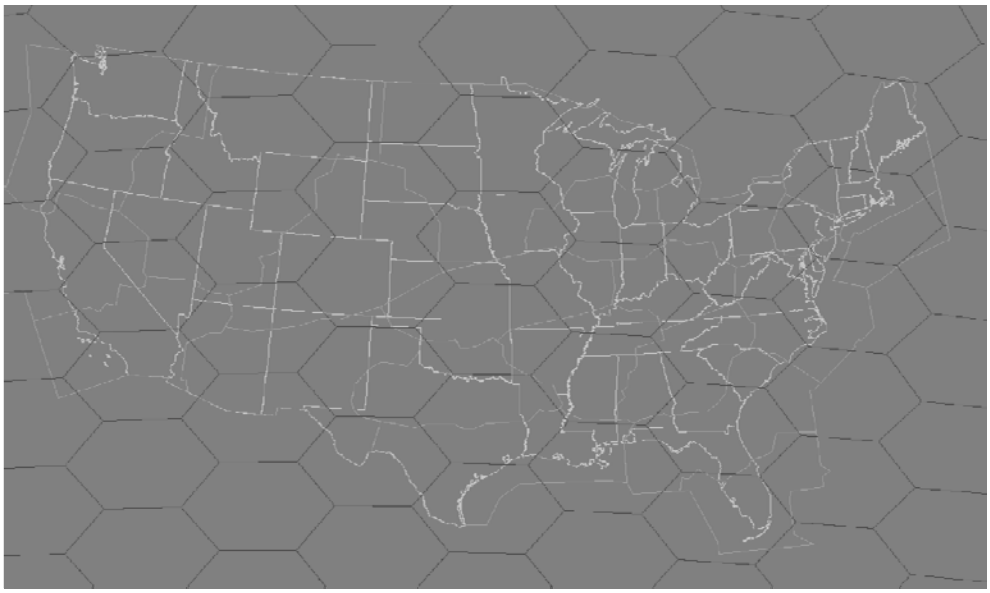

Figure 2.10: 300-nm hexagons over CONUS.

2.2.5 Verification at CCFP-like scale

Verifying products at spatial scales of the CCFP is an additional way to measure skill of convective products in an operational context. This work closely follows Lack et al. (2010a). Changing a deterministic observation or forecast field into CCFP-like scales requires the use of a clustering methodology. Fast Fourier Transform (FFT) band pass filters are used to convert spatial intensity to frequency space following Lack et al. (2010b). An example of this clustering technique is shown in **Figure 2.11** for radar reflectivity over Texas. For each cluster identified, the percent coverage of observations is calculated for those clusters exceeding minimum size criterion for CCFP (3000 mi^2). From the obtained percent coverage for the identified cluster, a

CCFP coverage category is assigned based on historical distributions of actual CCFP coverage for sparse/low, sparse/high and medium and above polygons (**Table 2.1**). An example clustering of the CIWS analysis field is shown in **Figure 2.12**. This example attempts to find areas of strong frequency signals above the VIP-level 2 threshold. In this case, the convection in N Alabama and W Tennessee does not get a CCFP polygon drawn around it due to its isolated nature and large separation between convective objects. Isolated convection typical of afternoon Gulf Coast storms will often have CCFP polygons drawn to the density of isolated convection. A clustering methodology that includes a series of erosion and dilation operations was also used in the analysis.

From these scaled images, skill scores such as CSI and bias can be calculated for a chosen coverage threshold matching the CCFP polygon forecast in the domain. This provides a useful baseline diagnostic comparison between the operational standard and the new forecast product being considered for traffic flow management. For the probabilistic LAMP forecast, thresholds are applied for particular issue/lead combinations so that the median bias is unity for the different CCFP coverage/confidence thresholds of interest. The table of calibrations for LAMP is shown in **Table 2.2** and is broken down by CONUS domain and NE domain.

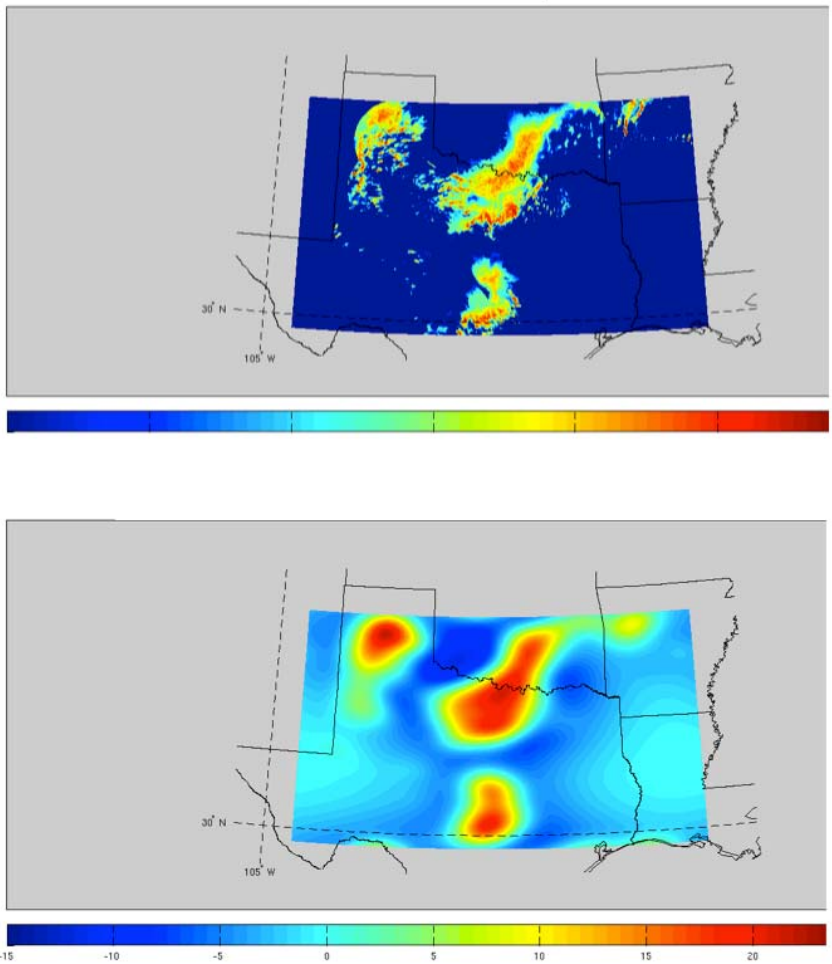

Figure 2.11: Scaling example for a particular convective event in Texas and Oklahoma. Actual reflectivity (top), with the output from the clustering algorithm (bottom) in arbitrary units, which can then be converted to areas assigned to coverage categories.

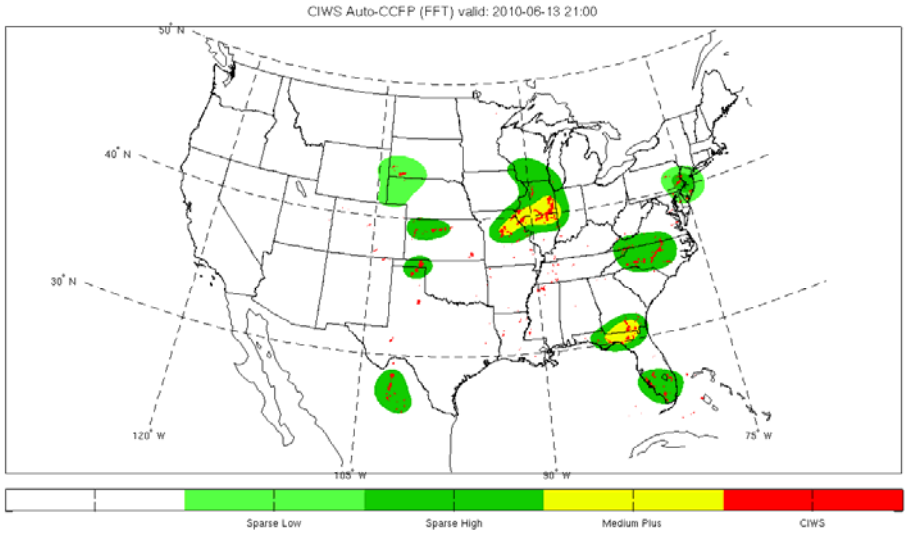

Figure 2.12: An example of a scaled CIWS field valid at 2100 UTC on 13 June 2010. Light green represents sparse/low coverage, dark green sparse/high coverage, yellow medium coverage or higher, and red is actual CIWS VIL at 3.5 kg m^{-2} or greater.

Table 2.2. LAMP probability (%) thresholds for the creation of CCFP-like polygons.

Lead 2	**CONUS**	**NE**
Sparse/Low	11	10
Sparse/High	16	14
Medium Plus	37	35
Lead 4	**CONUS**	**NE**
Sparse/Low	12	10
Sparse/High	15	13
Medium Plus	24	22
Lead 6	**CONUS**	**NE**
Sparse/Low	11	10
Sparse/High	14	12
Medium Plus	21	17
Lead 12	**CONUS**	**NE**
Sparse/Low	10	9
Sparse/High	13	11
Medium Plus	20	16
Lead 24	**CONUS**	**NE**
Sparse/Low	10	9
Sparse/High	12	11
Medium Plus	18	16

3. Analysis

3.1 Climatology

This section describes the character of the forecasts and observed weather during the study period, as a basis for the rest of the analysis.

Figure 3.1 shows the diurnal timing of convection in the two sub-regions. While the results differ, both indicate that National Airspace System (NAS) strategic planning times (nominally, 1100 through 1500 UTC) are scheduled before the initiation of convective weather that is of primary concern to planners. This study primarily focuses on forecasts issued before daily initiation, which are valuable to the planning process, but very challenging from a meteorological perspective. Much of the analysis focuses on 2100 UTC valid forecasts as representative of the performance of the product during difficult but important planning times.

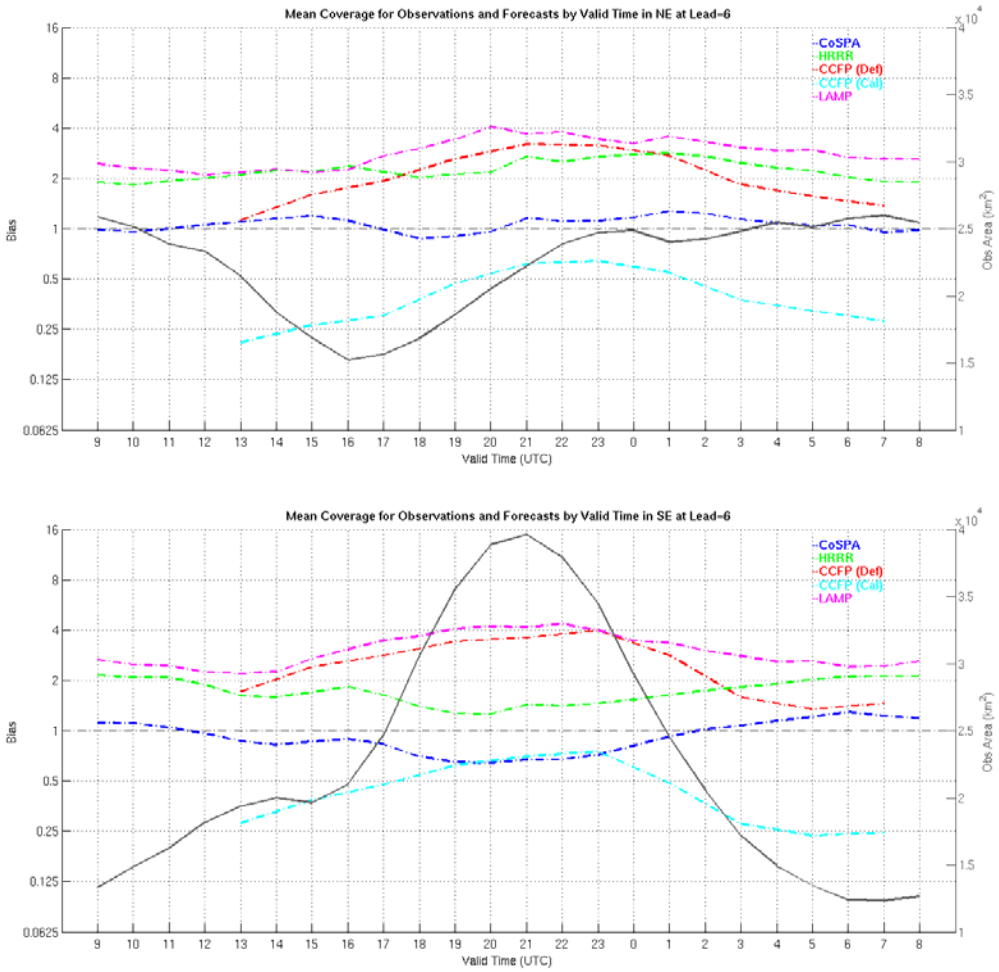

Figure 3.1: Convective coverage (VIP-level 3 or greater) by hour of day for the northeast (top) and the southeast (bottom). The solid curves represent the coverage of the observed weather (right y-axis), while the dotted lines represent bias for the different products at the 6-h lead-time (left y-axis). Bias is given on a log base 2 scale so that distance away from 1 is shown as equal penalty for over- or under-forecasting. At the time of strategic planning, both curves are near minimum.

Figure 3.2 shows CIWS-observed VIL at an equivalent VIP-level 3 threshold, valid each day at 2100 UTC and integrated over the entire study period. The geographic variation of convection is evident, highlighting the frequent occurrence of air mass storms in the southeast.

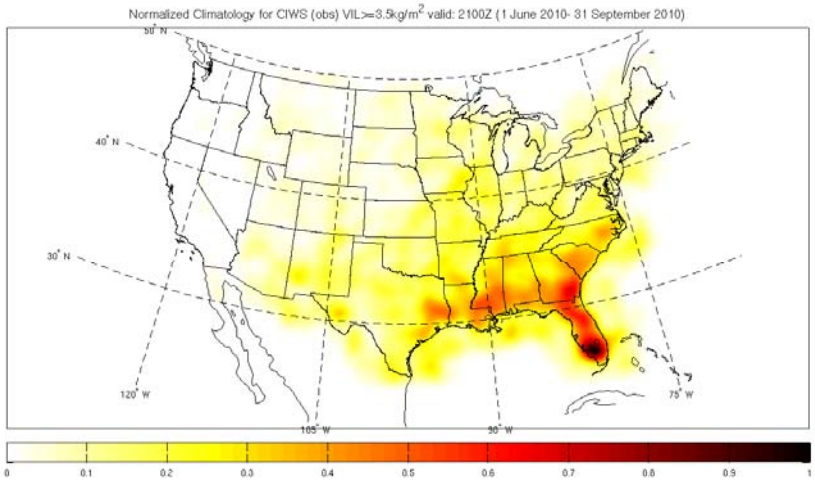

Figure 3.2: Normalized climatology of CIWS VIL analysis at a threshold of VIP-level 3 valid at 2100 UTC each day from 1 June 2010 through 30 September 2010. Larger values (darker colors) indicate higher occurrence of convective weather.

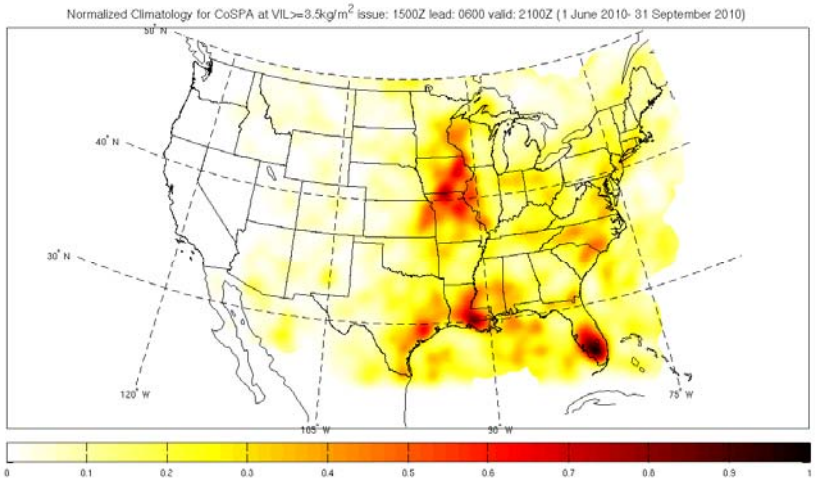

Figure 3.3: Normalized climatology of CoSPA forecast (1500 UTC issuance, 6-h lead-time) at a threshold of VIP-level 3 valid at 2100 UTC each day from 1 June 2010 through 30 September 2010. Larger values (darker colors) indicate higher occurrence of forecast convection.

Figure 3.3 shows the climatology of the CoSPA forecast during the study period. Overall, the forecast significantly under-forecasts convection in the southeast (compare **Figure 3.2** to **Figure 3.3**). Outside of the southeast, however, CoSPA exhibits structure on scales similar to that of the observed climatology.

The climatology of the LAMP Thunderstorm Product is depicted in **Figure 3.4**. Based heavily on climatology, the forecast shows broad, smooth areas, with little fine structure visible in the observed climatology (see **Figure 3.2**). LAMP does compare well with the observed climatology in the southeast, where air mass convection dominates.

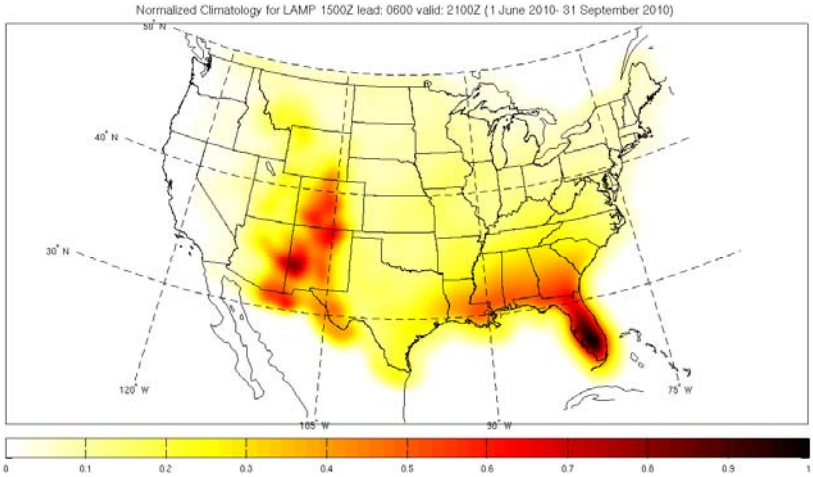

Figure 3.4: Normalized climatology of the LAMP Thunderstorm Product (1500 UTC issuance, 6-h lead-time) valid at 2100 UTC each day from 1 June 2010 through 30 September 2010. Larger values (darker colors) indicate higher occurrence of forecast convection.

Figure 3.5 shows the climatology of the calibrated CCFP forecast. The overall geographic pattern of forecast convection appears similar to that of the observations (compare **Figure 3.2** to **Figure 3.5**), with a coarser level of structure and a much higher frequency along the Gulf Coast.

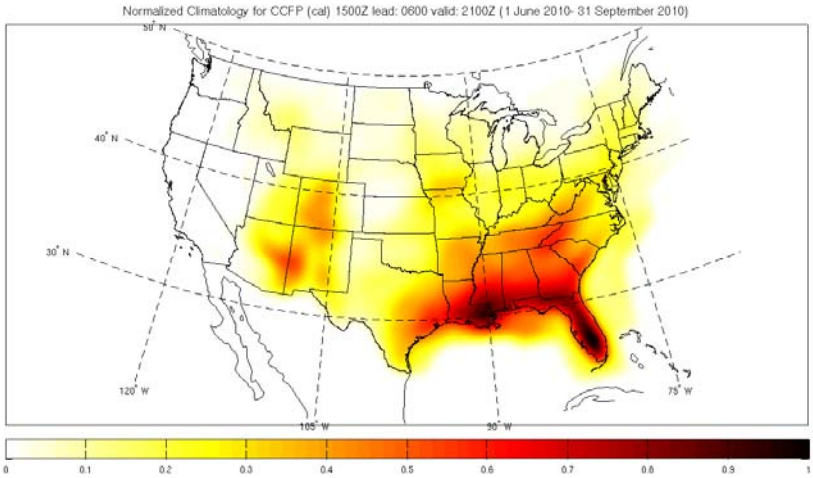

Figure 3.5: Normalized climatology of calibrated CCFP (1500 UTC issuance, 6-h lead-time) valid at 2100 UTC each day from 1 June 2010 through 30 September 2010. Larger values (darker colors) indicate higher occurrence of forecast convection.

3.2 Forecast performance at the 4 to 6-h lead-time

This section examines the performance of the forecasts during the 4 to 6-h lead-time, a period core to strategic planning for the morning hours.

3.2.1 Performance in the northeast domain

Figure 3.6 shows the FSS of forecasts versus resolution. By this measure, none of the forecasts perform very well at resolutions finer than about 45 km. CoSPA and calibrated CCFP outperform the other forecasts beginning at resolutions of about 70 km, the scale of a typical high-altitude sector. The coarse scale of information (smoothness) in LAMP and CCFP results in relatively flat FSS curves. All forecasts perform better than the uniform and climatological baselines of skill. The climatological skill is calculated by using the aggregate of occurrence of VIP-level 3 and greater convection during the 2010 convective study from the CIWS analysis field. At all resolutions, CoSPA scores better than the underlying HRRR used as input. Diagnostically, this seems to indicate that CoSPA blending does well in these circumstances. FSS plots with confidence intervals are included in the appendix of this report.

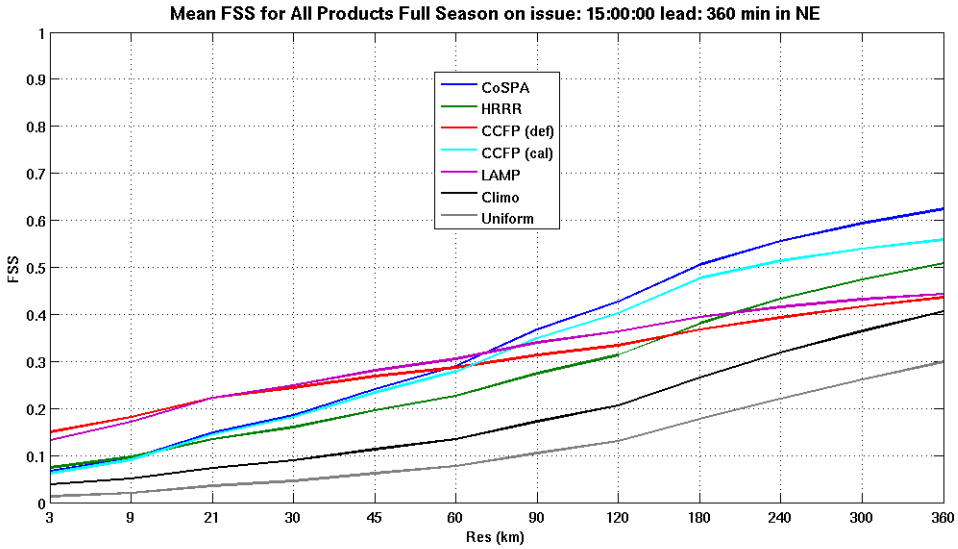

Figure 3.6: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), LAMP (purple), climo (black), and uniform (grey). Scores include northeast domain only, for products issued at 1500 UTC with a 6-h lead-time. Note: resolution is not a linear scale.

Plots of CSI versus Mincut threshold at 300-nm and 75-nm scales are shown in **Figure 3.7**. Overall, CoSPA and calibrated CCFP perform well. The structure of CoSPA and the underlying HRRR is evident in the higher CSI values for thresholds above 0.25. The low CSI values for LAMP and CCFP with increasing threshold indicates less sharpness in those forecasts. Corresponding bias plots of Mincut are shown in **Figure 3.8**. Mincut CSI plots with confidence intervals are included in the appendix of this report.

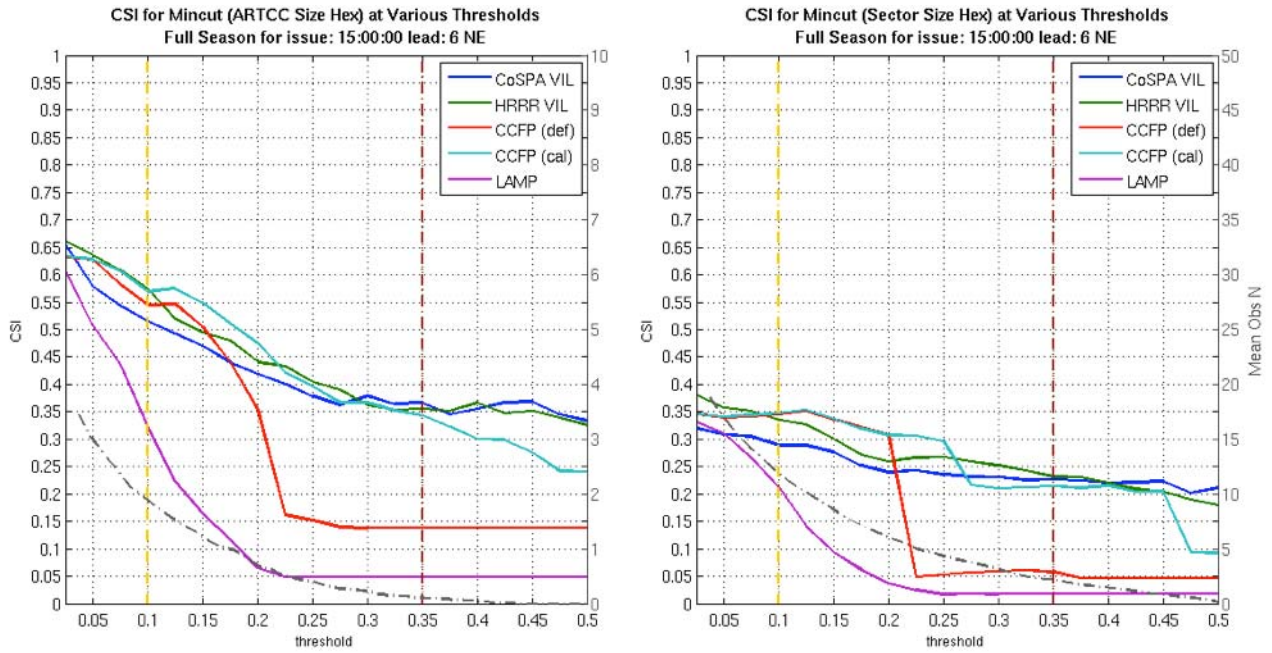

Figure 3.7: Critical Success Index (CSI) versus Mincut threshold for ARTCC scale (left) and sector scale (right). Products shown are CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), and LAMP (purple). The dashed grey line indicates the mean number of hexagons that are impacted at the given Mincut threshold for all days in the observed field, with the scale on the right of the plot. Vertical dashed lines represent medium impact events (yellow) and high impact events (maroon). Scores include northeast domain only, for products issued at 1500 UTC with a 6-h lead-time.

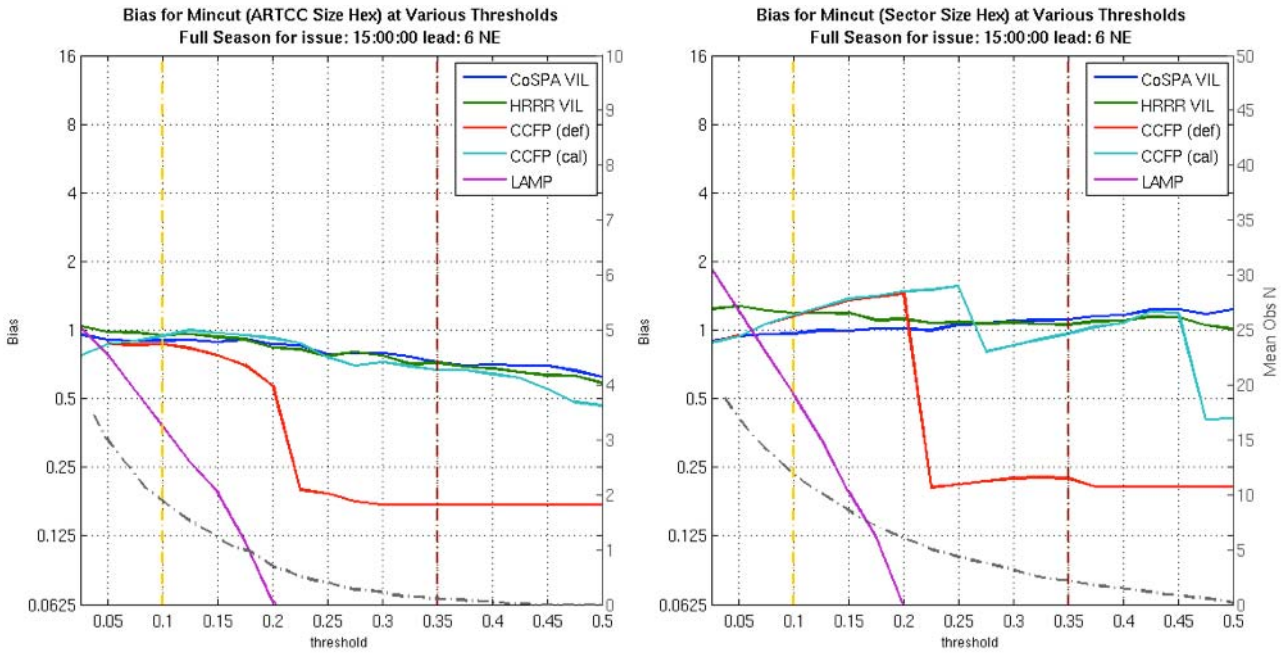

Figure 3.8: Bias versus Mincut threshold for ARTCC scale (left) and sector scale (right). Products shown are CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP(cyan), and LAMP (purple). The dashed grey line indicates the mean number of hexagons that are impacted at the given Mincut threshold for all days in the observed field, with the scale on the right of the plot. Vertical dashed lines represent medium impact events (yellow) and high impact events (maroon). Scores include northeast domain only, for products issued at 1500 UTC with a 6-h lead-time.

3.2.2 Performance in the southeast domain

Plots of FSS versus resolution for the southeast domain are shown in **Figure 3.9**. At scales more coarse than 40 km, calibrated CCFP outperforms the other products. For resolution up to 100 km, CoSPA scores low, only showing comparable skill to LAMP and CCFP at very coarse scales. The HRRR performs slightly better than CoSPA for all resolutions, diagnostically indicating that the underlying model might have a better bias in this region for the 6-h lead-time. Only calibrated CCFP and HRRR score better than the climatological baseline score at coarse resolutions in the air mass dominated southeast.

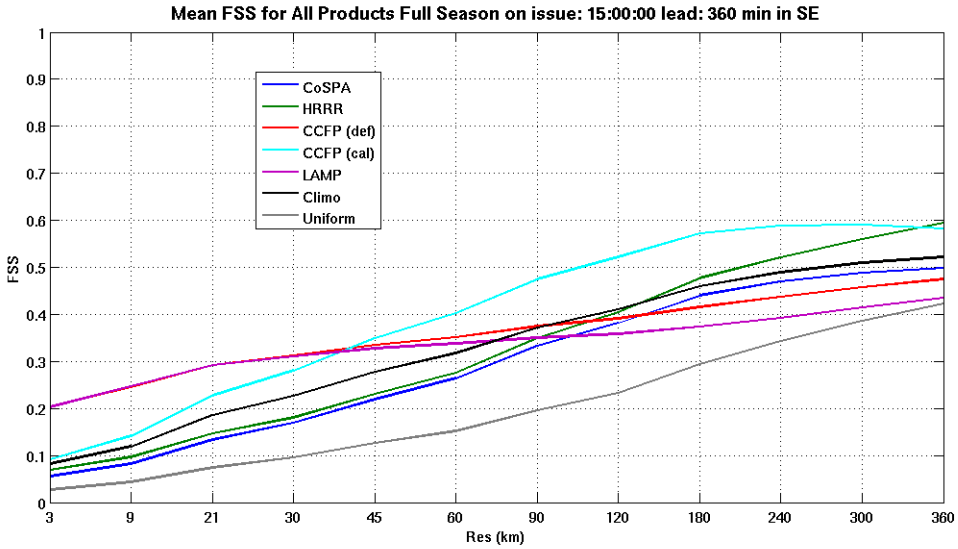

Figure 3.9: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), LAMP (purple), climo (black), and uniform (grey). Scores include southeast domain only, for products issued at 1500 UTC with a 6-h lead-time. Note: resolution is not a linear scale.

Figure 3.10 shows plots of CSI versus Mincut threshold at ARTCC and sector scale in the southeast. For low and moderate thresholds at this scale, LAMP and Calibrated CCFP show good skill, while CoSPA has moderate skill. In the Mincut plot, CoSPA starts to outperform HRRR at more impactful thresholds unlike the FSS. However, for the Mincut scores, all products are calibrated, which shows that with additional calibration CoSPA may gain back the apparent loss in skill from the FSS metric. Corresponding bias plots are shown in **Figure 3.11**.

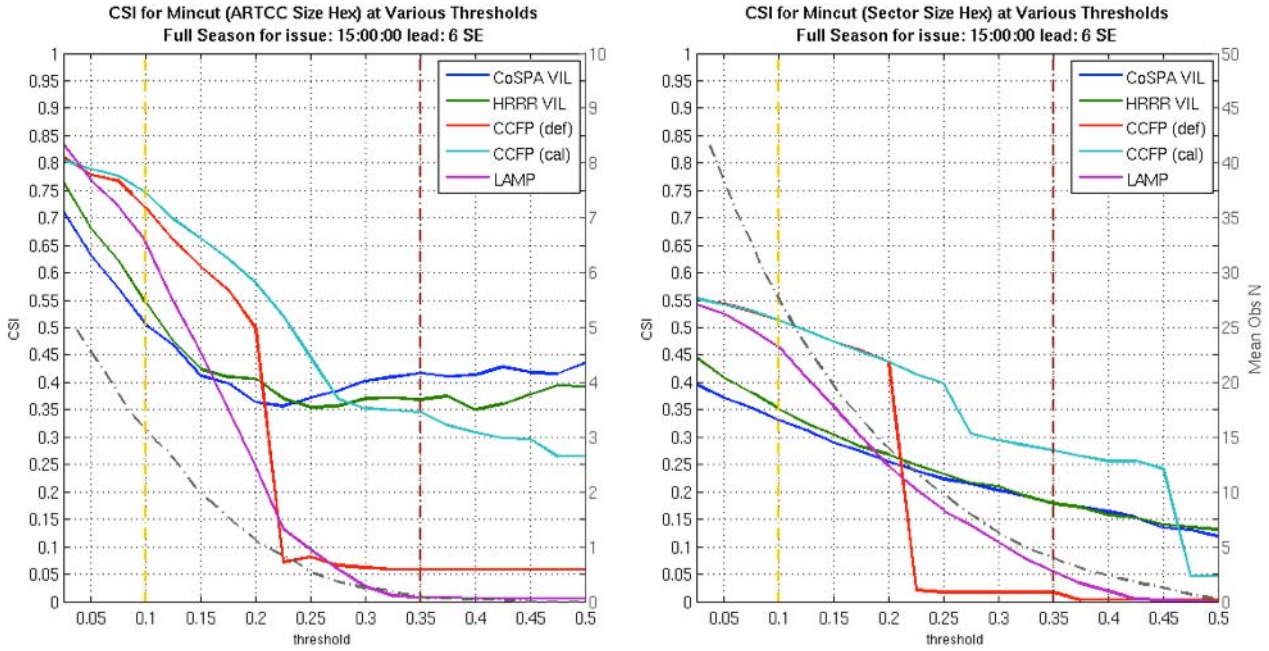

Figure 3.10: Critical Success Index (CSI) versus Mincut threshold for ARTCC scale (left) and sector scale (right). Products shown are CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), and LAMP (purple). The dashed grey line indicates the mean number of hexagons that are impacted at the given Mincut threshold for all days in the observed field, with the scale on the right of the plot. Vertical dashed lines represent medium impact events (yellow) and high impact events (maroon). Scores include southeast domain only, for products issued at 1500 UTC with a 6-h lead-time.

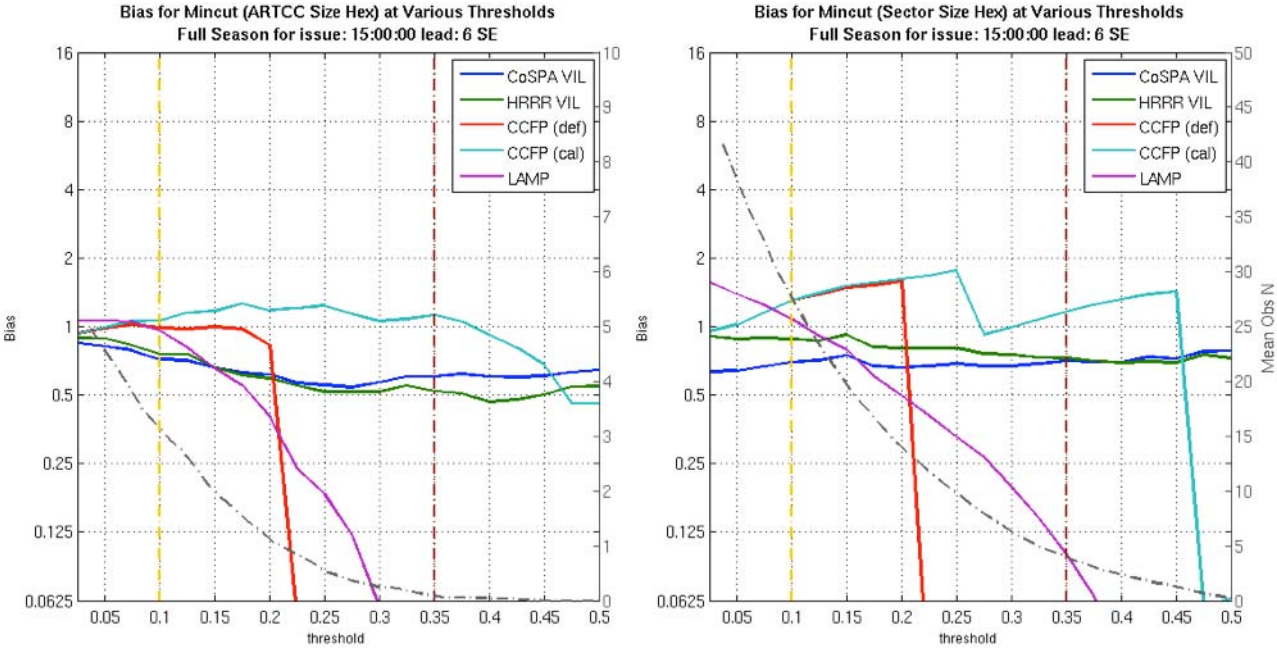

Figure 3.11: Bias versus Mincut threshold for ARTCC scale (left) and sector scale (right). Products shown are CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), and LAMP (purple). The dashed grey line indicates the mean number of hexagons that are impacted at the given Mincut threshold for all days in the observed field, with the scale on the right of the plot. Vertical dashed lines represent medium impact events (yellow) and high impact events (maroon). Scores include southeast domain only, for products issued at 1500 UTC with a 6-h lead-time.

3.3 Forecast performance at the 2 to 4-h lead-time

This section focuses on the performance of forecasts at the 2 to 4-h lead-time. First, all of the products are measured at the 2-h lead-time, and then CoSPA performance is examined beginning at the 2:15 lead-time, where the forecast algorithm differs from CIWS.

3.3.1 Performance in the northeast domain

A plot of FSS versus resolution for all products in the northeast domain at the 2-h lead-time is shown in **Figure 3.12**. By this measure, CoSPA clearly outperforms the other forecasts, showing good skill at resolutions down to about 15 km, a scale much finer than that of the high-altitude sectors.

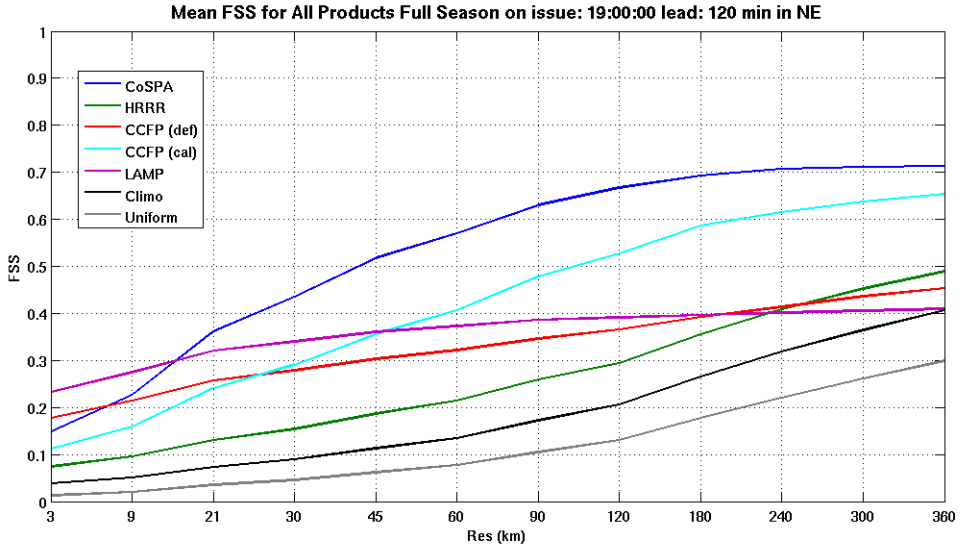

Figure 3.12: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), LAMP (purple), climo (black), and uniform (grey). Scores include northeast domain only, for products issued at 1900 UTC with a 2-h lead-time. Note: resolution is not a linear scale.

3.3.2 Performance in the southeast domain

Figure 3.13 shows FSS versus resolution in the southeast domain. CoSPA's characteristics in the southeast are similar to those in the northeast. Again, it significantly outperforms the other forecasts, with the exception of calibrated CCFP, which in this domain at the 2-h lead-time performs comparably.

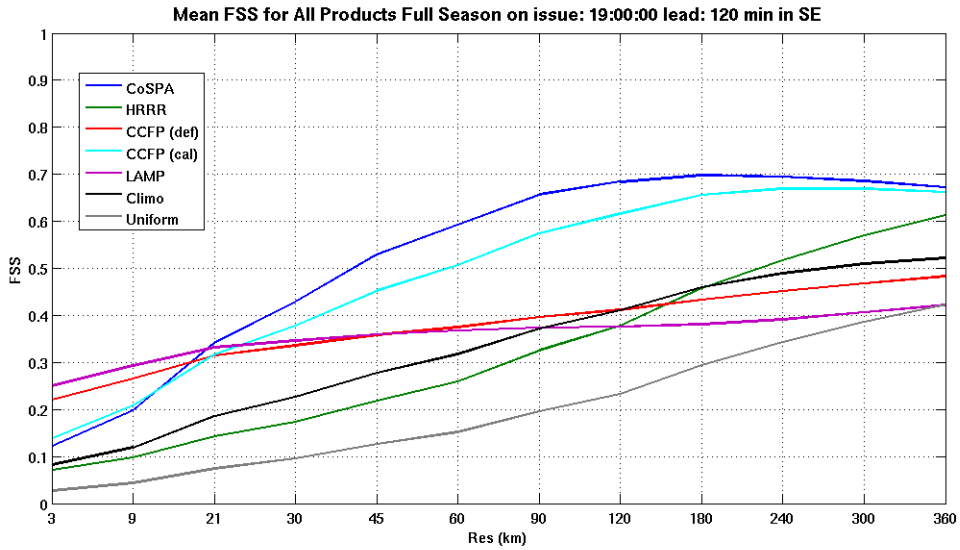

Figure 3.13: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), LAMP (purple), climo (black), and uniform (grey). Scores include southeast domain only, for products issued at 1900 UTC with a 2-h lead-time. Note: resolution is not a linear scale.

Although the 2-h lead-time issued at 1900 UTC performs well, 2100 UTC is the height of convective activity in the southeast domain. At issuances before convective initiation, there is a notable dip in the bias shown in **Figure 3.14**. During convective initiation times (1600-2000 UTC) CoSPA issuances, on average, have a bias of about 0.5.

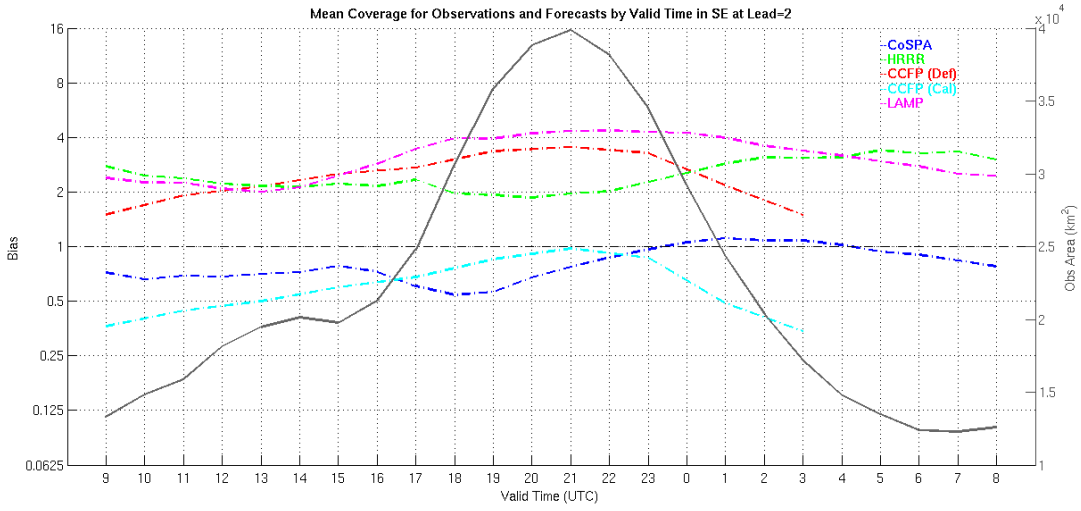

Figure 3.14: Convective coverage (VIP-level 3 or greater) by hour of day for the southeast domain. The solid curve represents the coverage of the observed weather (right y-axis), while the dotted lines represent bias for the different products at the 2-h lead-time (left y-axis). Bias is given on a log base 2 scale so that distance away from 1 is shown as equal penalty for over- or under-forecasting.

3.3.3 CoSPA performance for 2 to 4-h lead-times

This section examines the performance of CoSPA for each of the 15-min lead intervals between the 2-h and 4-h lead-times. Overall, FSS results in **Figure 3.15** indicate that the forecast is well behaved during this interval; that is, the FSS values decrease reasonably as lead-time increases. The nature of the curve for the 2:00 lead-time (blue) appears to differ from the "family" of other lead-time curves. This difference might result from the discontinuity of the CoSPA forecast at 2:00 and 2:15, which has been observed subjectively in many instances by animating images through the time interval.

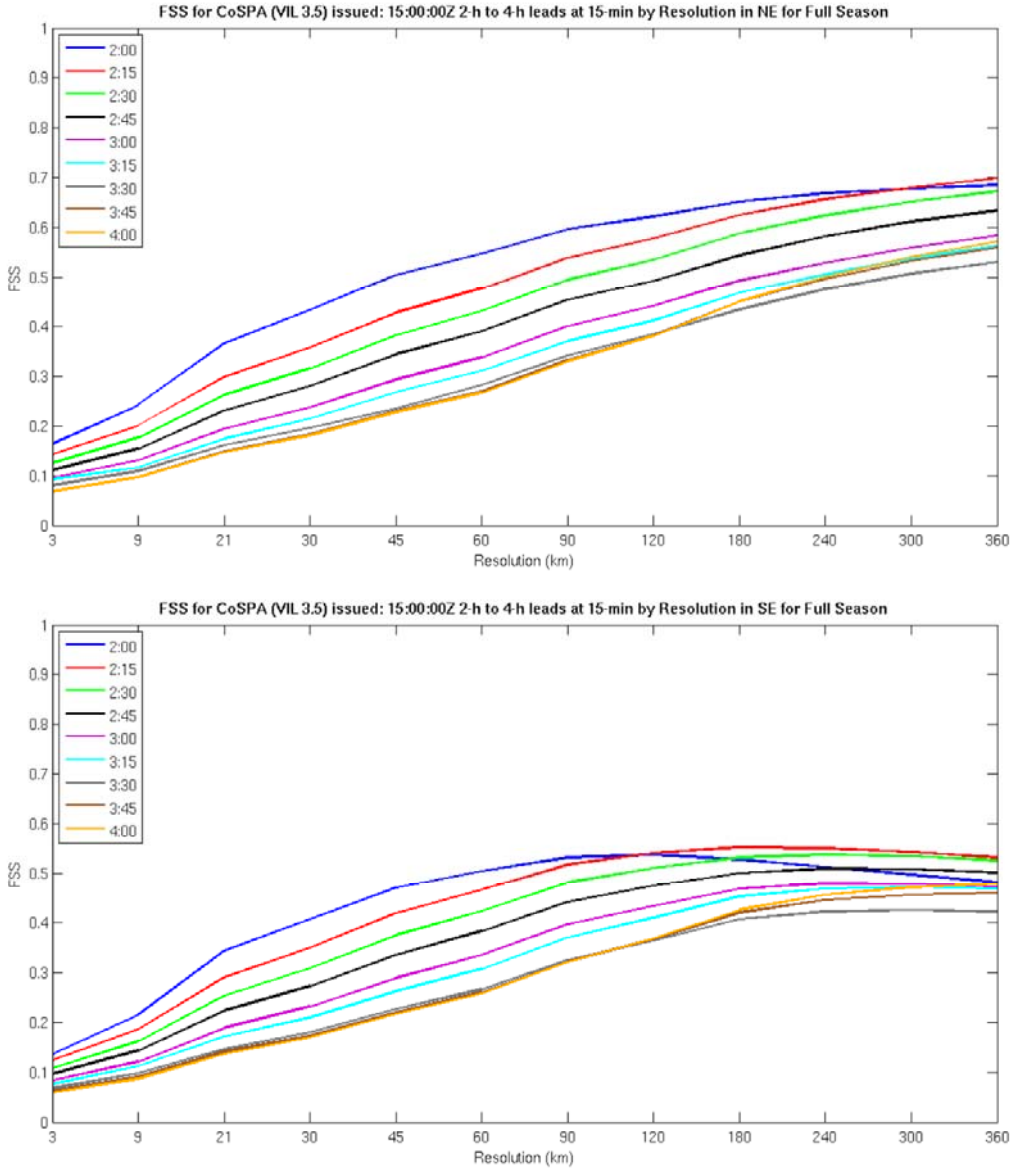

Figure 3.15: Fractions Skill Score versus resolution for CoSPA lead times between two and four hours in the northeast domain (top) and the southeast domain (bottom), for products issued at 1500 UTC. Note: resolution is not a linear scale.

3.4 Performance at the 8-h lead-time

This section examines the performance of products at the 8-h lead-time, a CCFP outlook that may become available for the 2011 convective season.

Figure 3.16 shows plots of FSS versus resolution for the three products that issue forecasts with an 8-h lead-time: CoSPA, HRRR, and LAMP. Overall, these products perform similarly to their 6-h lead-time counterparts. The relative observations about the 6-h lead products hold for the 8-h forecasts in the northeast domain; however, a forecast based on climatology has skill greater than most of the products in the southeast domain.

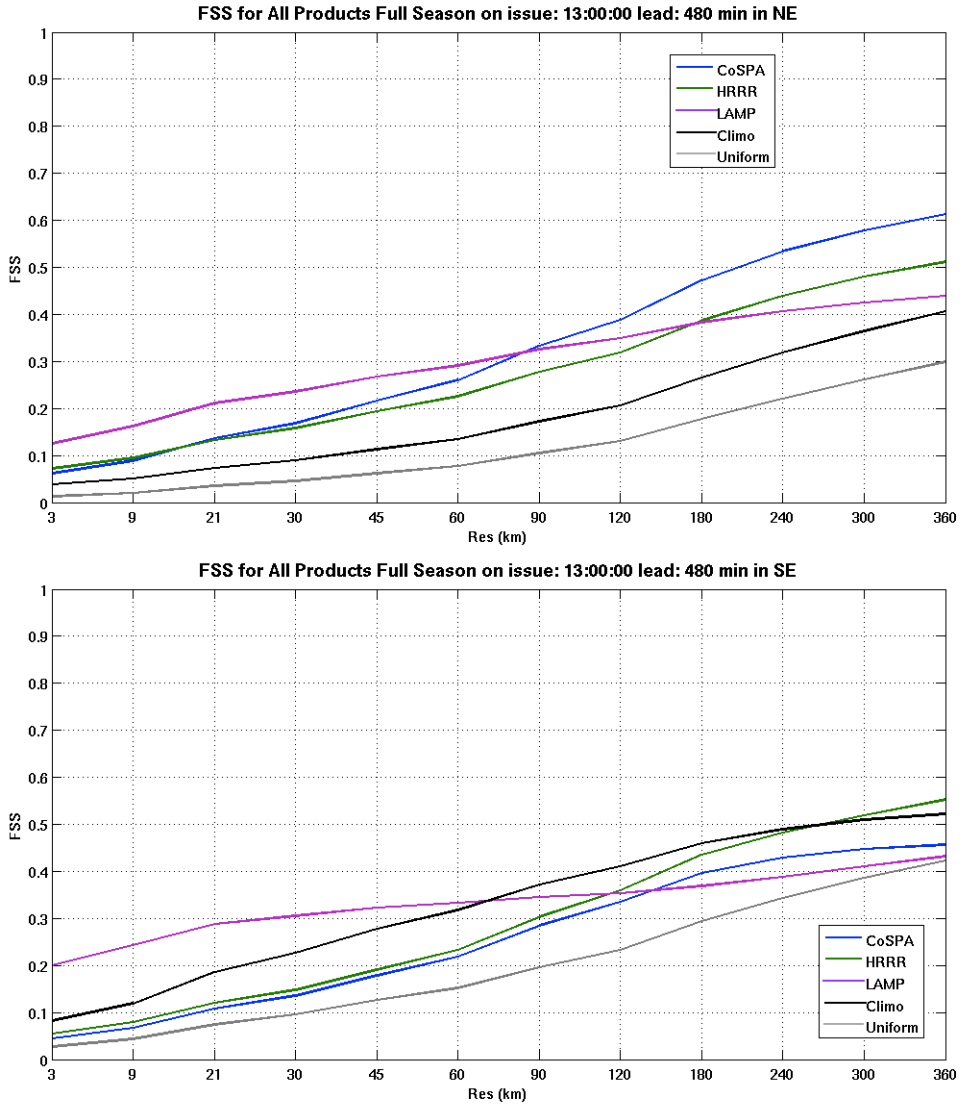

Figure 3.16: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), LAMP (purple), climo (black) and uniform (grey) in the northeast domain (top) and the southeast domain (bottom), for products issued at 1300 UTC with an 8-h lead-time. Note: resolution is not a linear scale.

3.5 Supplemental relationship to CCFP

3.5.1 CCFP sparse coverage/low confidence polygons

During the study period, at the 6-h lead-time during telecon issuances (1100, 1300, and 1500 UTC), the CCFP forecast included approximately 1100 polygons of the sparse coverage/low confidence type. Of these polygons, 165 contained observed coverage at the medium level (yellow in **Figure 2.12**), as measured by the scaling technique. In these cases, which likely include weather of concern to planners, CoSPA forecast areas of medium level coverage with a PODy of 0.65. In the cases where the polygons did not contain any observed coverage at the medium level, CoSPA identified these situations with a PODn of 0.81. Thus, CoSPA might serve as a discriminator of CCFP sparse coverage/low confidence polygons in morning issuances for air traffic managers. LAMP appears as though it does not effectively discriminate between polygons with and without areas of medium level coverage (PODy of 0.53 and a PODn of 0.46).

3.5.2 CoSPA/CCFP relationship

As part of the analysis, a supplemental decomposition was performed to determine areas of agreement between CoSPA and CCFP. Within the areas of agreement, CoSPA performed substantially better than it did when disagreeing with CCFP, as measured by FSS over many resolutions. The peak of the FSS curve (1500 UTC issuance, 6-h lead-time) when in agreement is ~0.4, while the peak when disagreeing is ~0.05 (figure not shown).

3.5.3 Long-range planning

Figure 3.17 is a plot of CSI versus lead-time for the LAMP forecast, as measured by the scaling technique used to identify areas of sparse convective coverage. The forecasts for 12-h and 24-h lead-times perform comparably to those of the 6-h and 8-h lead-times. This suggests that LAMP provides useful information for next-day planning, for convective coverage at the broadest scale.

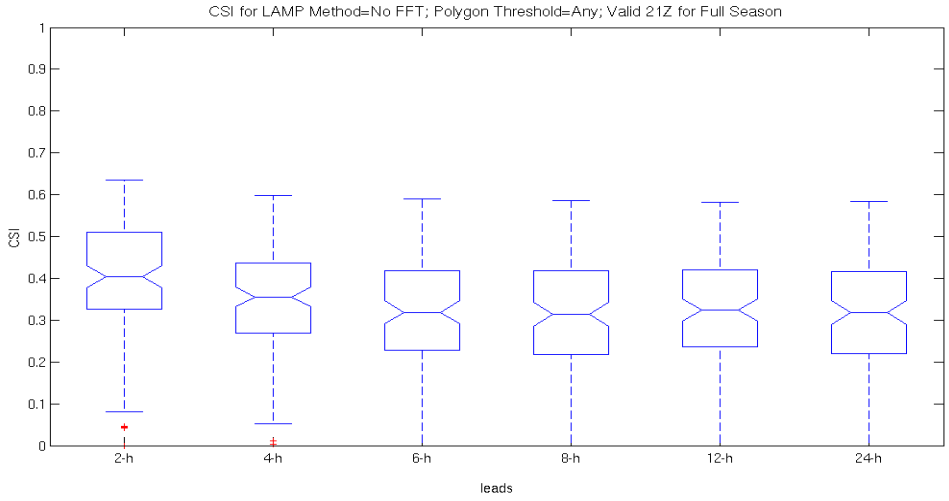

Figure 3.17: CSI versus lead-time for LAMP, each forecast lead time valid at 2100 UTC, as determined by the sparse category of the scaling technique.

3.6 Blitz Days versus Full Season Skill

During the evaluation period, the Operational Readiness and Impact Team within FAA's Aviation Weather Group selected several days, designated as "blitz days", for field evaluations of both the CoSPA and LCH products. Requiring notice for travel logistics, the team selected the blitz days in advance. Blitz days, in general, involved convection resulting from a strongly forced synoptic environment. This section is included in the analysis to compare the climatology of the blitz days to the overall season, and to indicate differences in forecast performance between the two periods.

3.6.1 Climatology of Blitz Days

Figure 3.18 shows a comparison of climatology for "blitz days" versus that of the full convective season. During the blitz days, most of the convection occurs along frontal boundaries extending from New England to Texas. In the full-season climatology it is evident that the signal is dominated by air mass type thunderstorms in the southeast.

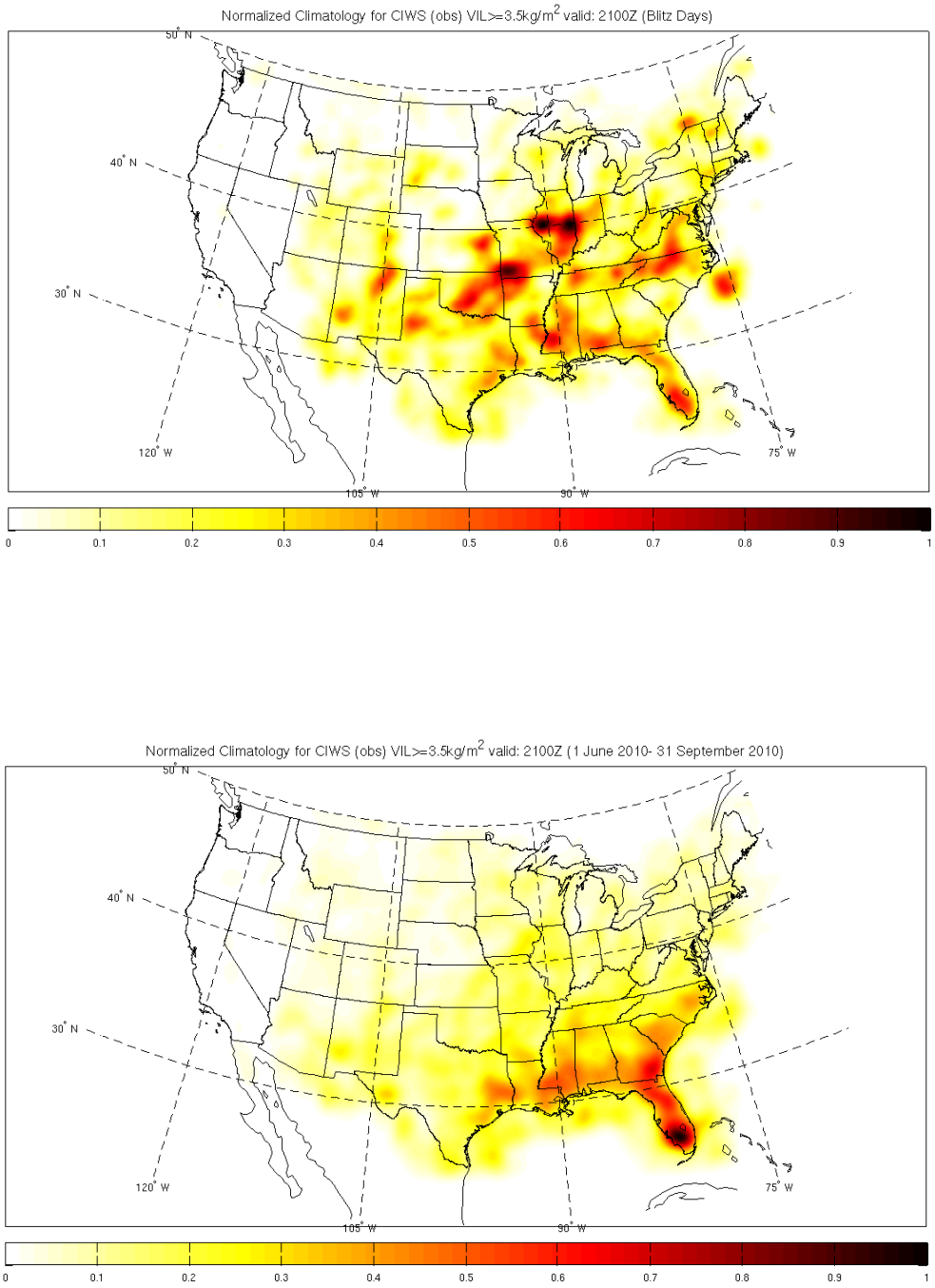

Figure 3.18: CIWS climatology for the blitz days selected for human evaluation (top) compared with the CIWS climatology for the full season (bottom) valid at 2100 UTC.

3.6.2 Blitz vs. Full Season Skill Comparison

Plots of the skill of the forecast products show higher performance on "blitz days", likely due to stronger synoptic forcing present during those days. **Figure 3.19** shows the comparison between FSS for the full season and that of the "blitz days" in the northeast domain. For CoSPA, the FSS is near 0.5 around 90 km for "blitz days" while near 0.35 around 90 km for all days in the study period. The skill in the southeast domain is roughly the same for both "blitz days" and the full season (figure not shown), likely due to the air mass nature of convection dominant in the region.

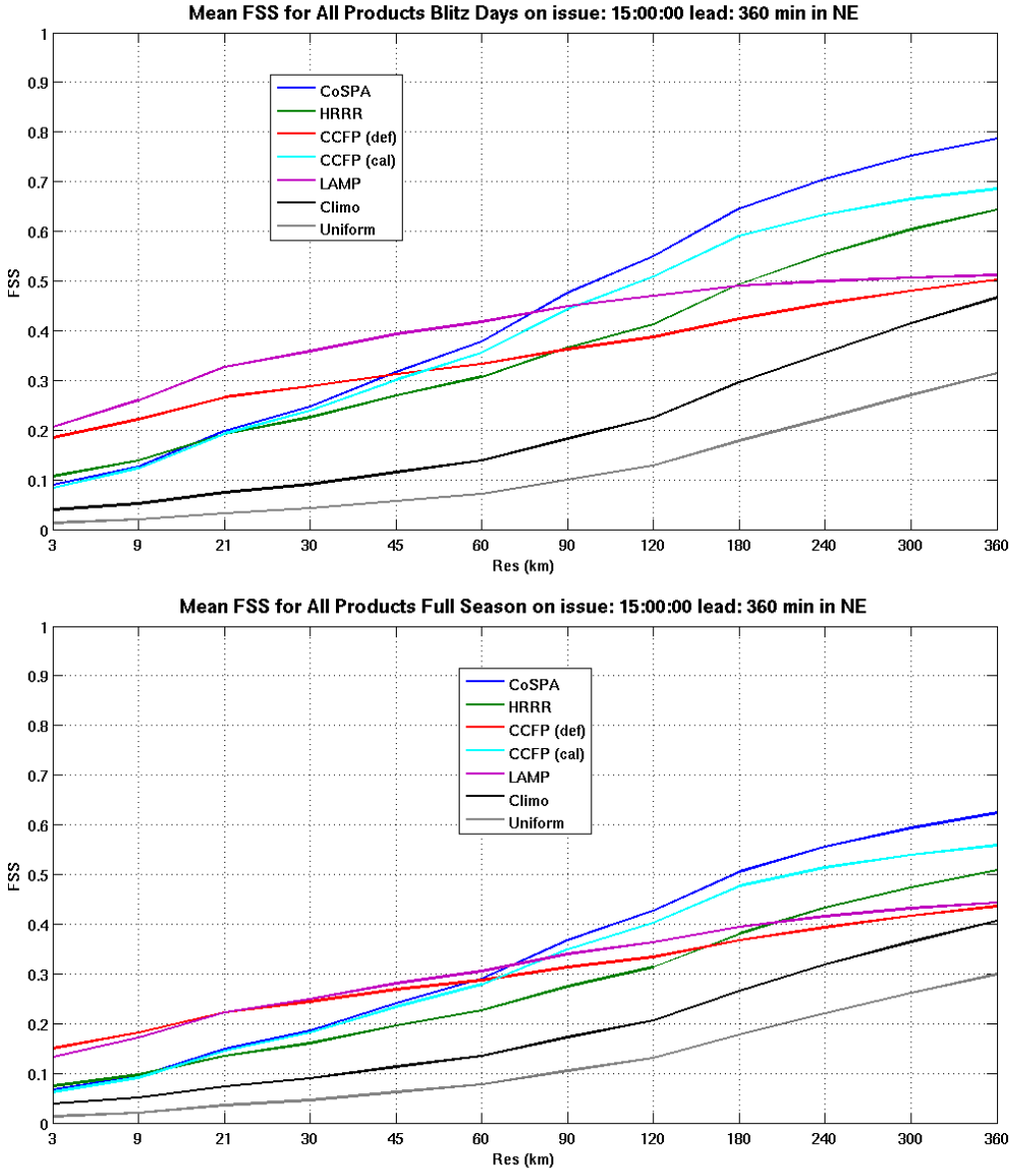

Figure 3.19: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), LAMP (purple), CCFP (red), calibrated CCFP (cyan), climo (black) and uniform (grey) in the northeast domain for blitz days (top) and for the full season (bottom), for products issued at 1500 UTC with a 6-h lead-time. Note: resolution is not a linear scale.

4. Conclusions

This report summarizes a formal quality assessment of CoSPA, a high-resolution convective forecast being developed for use in air traffic management. On behalf of the Federal Aviation Administration's Aviation Weather Research Program, and in support of an Aviation Weather Technology Transfer (AWTT) D4 (operational) decision point, this study was carried out during the 2010 convective season as outlined in the FAA Aviation Weather Group's *Comprehensive Plan for the CoSPA 2010 Demonstration*.

Findings from this report are as follows:

Climatology

This analysis determined that convective weather regimes differ significantly in the northeast and southeast U.S., with corresponding delineation in forecast quality in these two domains.

Performance at the 4 to 6-h lead-time

In the northeast, CoSPA outperforms the other forecasts, showing good skill on a scale similar to that of high-altitude sectors used for TFM. Weather translation metrics indicate that CoSPA forecasts more intense, rare events (sharpness), while LAMP does not.

In the southeast, weather metrics indicate LAMP and calibrated CCFP perform significantly better than CoSPA, but only at relatively coarse scales. Weather translation metrics indicate that LAMP only forecasts lower-intensity events at ARTCC-scale, while calibrated CCFP appears to provide information about the higher intensity ARTCC-scale events.

Performance at the 2 to 4-h lead-time

In the northeast and southeast, both weather and weather translation metrics indicate that CoSPA performs as well as or better than the other forecasts. CoSPA does exhibit a visible discontinuity between the 2:00 (actually the CIWS 2:00) and the 2:15 leads, likely due to the difference between CIWS and the underlying blended forecast. Beyond that time, CoSPA forecasts appear to be consistent.

Performance at the 8-h lead-time

Overall, the 8-h lead-time forecasts of CoSPA and LAMP perform similar to their 6-h lead-time counterparts. The observations about the products at the 6-h lead-time pertain also to the 8-h lead-time. Thus, the 8-h lead-time forecasts appear to have planning value similar to that of the 6-h forecasts.

Supplemental relationship to CCFP

Analysis based on the use of the forecasts in conjunction with one another indicates:

- CoSPA adds information to CCFP sparse coverage, low confidence polygons issued at strategic planning times, by effectively discriminating those that will contain significant areas of convection.
- CoSPA performs much better inside CCFP polygons than in areas outside these polygons. Planners might use the finer structure provided by CoSPA inside the polygons with a greater degree of confidence.
- LAMP provides useful information for next-day planning (12 and 24-h lead-times) for convective coverage at the broadest scale.

Acknowledgements

This research is in response to requirements and funding provided by the Federal Aviation Administration. The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.

5. References

Dupree, W., J. Pinto, M. Wolfson, S. Benjamin, M. Steiner, J. Williams, D. Morse, X. Tao, D. Ahijevych, H. Iskenderian, C. Reiche, J. Pelagatti, and M. Matthews, 2009: The Advanced Storm Prediction for Aviation Forecast Demonstration, *WMO Symposium on Nowcasting, Whistler, B.C., Canada.*

Ghirardelli, J., 2005: An overview of the redeveloped Localized Area MOS Program (LAMP) for short-range forecasting. *Preprints, 21st Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction, Washington, D.C.*

Krozel, J., S. Penny, J. Prete, and J. S. B. Mitchell, 2004: Comparison of Algorithms for Synthesizing Weather Avoidance Routes in Transition Airspace, *AIAA Guidance, Navigation, and Control Conf., Providence, RI.*

Lack, S.A., G.J. Layne, M.P. Kay, S. Madine, and J. Mahoney, 2010a: Evaluating supplemental relationships between convective forecast products for aviation. *Preprints, 14th Conference on Aviation, Range, and Aerospace Meteorology (ARAM). Atlanta, GA.*

Lack, S. A., G. L. Limpert, and N. I. Fox, 2010b: An object-oriented multi-scale verification scheme. *Weather and Forecasting*, **25**, 79-92.

Layne, G.J. and S.A. Lack, 2010: Methods for estimating air traffic capacity reductions due to convective weather for verification. *Preprints, 14th Conference on Aviation, Range, and Aerospace Meteorology (ARAM). Atlanta, GA.*

Phillips, C., J. Pinto, D. Albo, and M. Steiner, 2010: Diagnosis of phase errors in high-resolution NWP model forecasts of precipitation and application of improved aviation weather forecasts, *Preprints, 14th Conference on Aviation, Range, and Aerospace Meteorology (ARAM), Atlanta, GA.*

Roberts, N. M. and H. W. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Mon. Wea. Rev.*, **136**, 78–97.

Weygandt, S., T. Smirnova, C. Alexander, S. Benjamin, K. Brundage, B. Jamison, and S. Sahm, 2010: The High Resolution Rapid Refresh (HRRR): Recent enhancements and evaluation during the 2009 convective season. *Preprints, 14th Conference on Aviation, Range and Aerospace Meteorology (ARAM), Atlanta, GA.*

Wolfson, M., W. Dupree, R. Rasmussen, M. Steiner, S. Benjamin, and S. Weygandt, 2008: Consolidated Storm Prediction for Aviation (CoSPA), *Preprints, 13th Conference on Aviation, Range, and Aerospace Meteorology (ARAM), New Orleans, LA.*

6. Appendix

6.1 FSS with Confidence Intervals

Confidence intervals for the mean fractions skill score plots above were also calculated. The bias corrected and accelerated (BCa) bootstrap technique was used for the 95% confidence interval. 10000 bootstrap samples were used in the calculations. The plot is shown in **Figure 6.** for the 1500 UTC issue 6-h lead-time.

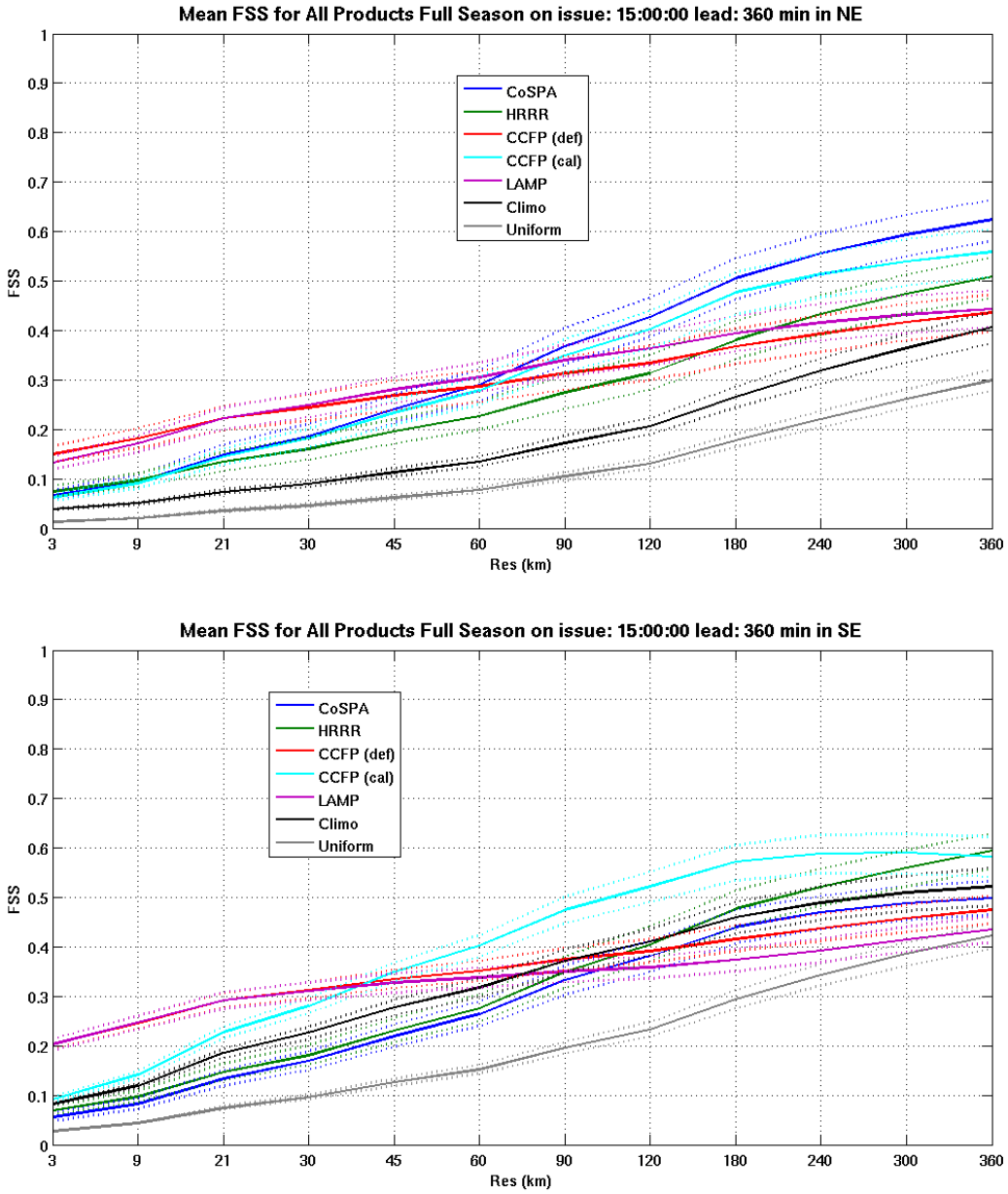

Figure 6.1: Fractions Skill Score versus resolution for CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), LAMP (purple), climo (black), and uniform (grey). Scores include northeast domain (top) and southeast domain (bottom), for products issued at 1500 UTC with a 6-h lead time. 95% confidence intervals are dotted. Note: resolution is not a linear scale.

6.2 Mincut with Confidence Intervals

Confidence intervals for the mean Mincut CSI plots above were also calculated. The bias corrected and accelerated (BCa) bootstrap technique was used for the 95% confidence interval. 10000 bootstrap samples were used in the calculations. The northeast plot for ARTCC and sector scale is shown in **Figure 6.2**, while the southeast plot for ARTCC and sector scale is shown in **Figure 6.3**.

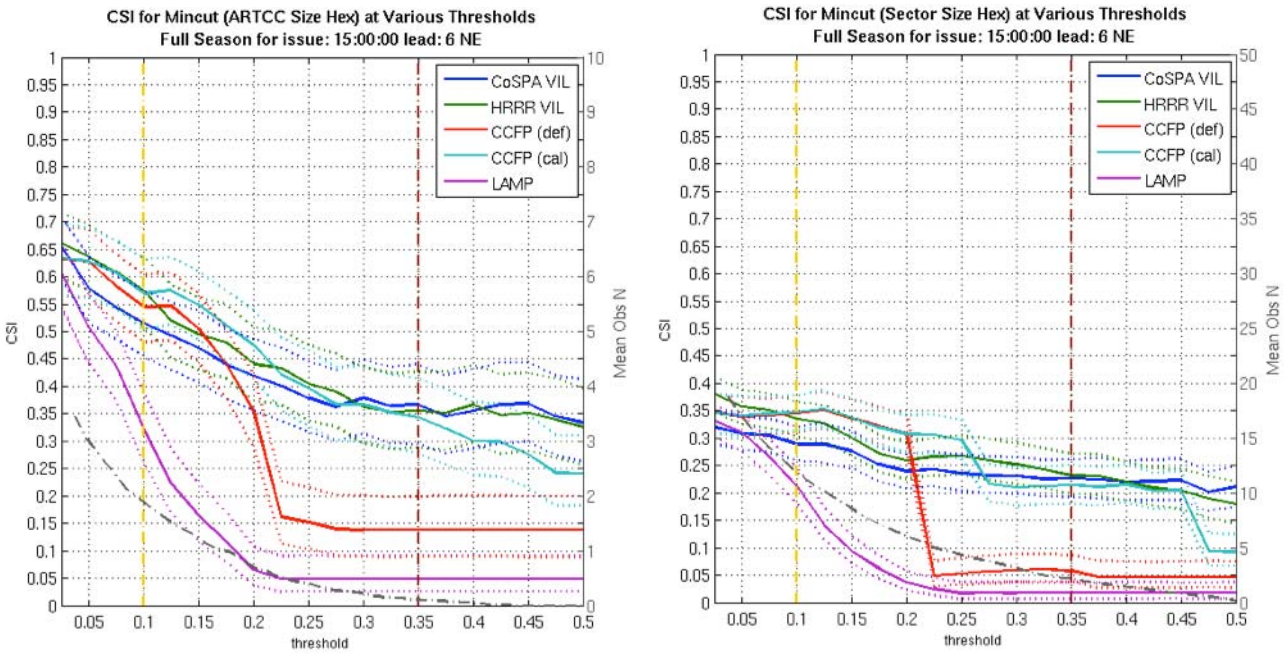

Figure 6.2: Critical Success Index (CSI) versus Mincut threshold for ARTCC scale (left) and sector scale (right). Products shown are CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), and LAMP (purple). The dashed grey line indicates the mean number of hexagons that are impacted at the given Mincut threshold for all days in the observed field, with the scale on the right of the plot. Vertical dashed lines represent medium impact events (yellow) and high impact events (maroon). Scores include northeast domain only, for products issued at 1500 UTC with a 6-h lead-time. 95% confidence intervals are dotted.

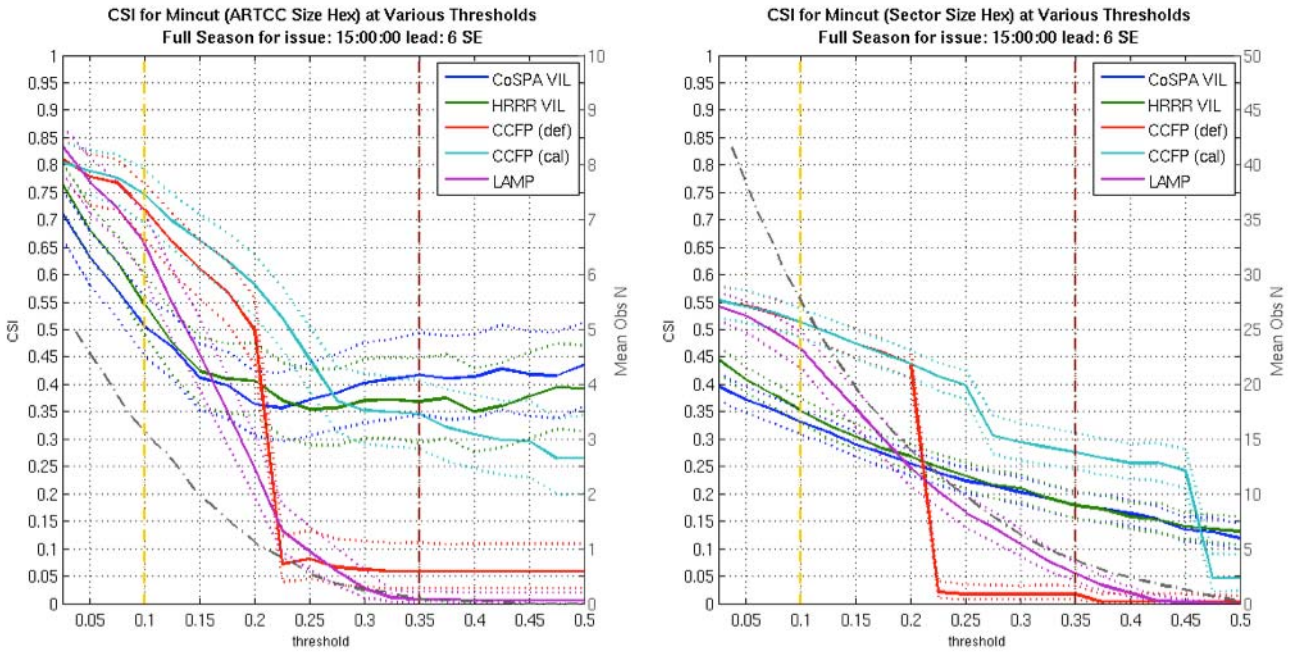

Figure 6.3: Critical Success Index (CSI) versus Mincut threshold for ARTCC scale (left) and sector scale (right). Products shown are CoSPA (blue), HRRR (green), CCFP (red), calibrated CCFP (cyan), and LAMP (purple). The dashed grey line indicates the mean number of hexagons that are impacted at the given Mincut threshold for all days in the observed field, with the scale on the right of the plot. Vertical dashed lines represent medium impact events (yellow) and high impact events (maroon). Scores include southeast domain only, for products issued at 1500 UTC with a 6-h lead-time. 95% confidence intervals are dotted.